Estimation and Inference in Games of Incomplete Information with Unobserved Heterogeneity and Large State Space

Yanqin Fan* Shuo Jiang† Xuetao Shi‡

July 28, 2024

Abstract

Building on the sequential identification result of Aguirregabiria and Mira (2019), this paper develops estimation and inference procedures for static games of incomplete information with payoff-relevant unobserved heterogeneity and multiple equilibria. With payoff-relevant unobserved heterogeneity, sequential estimation and inference face two main challenges: *the matching-types problem* and *a large number of matchings*. We tackle the matching-types problem by constructing a new minimum-distance criterion for the correct matching and the payoff function with both correct and incorrect "moments". To handle large numbers of matchings, we propose a novel and computationally fast multistep moment selection procedure. We show that asymptotically, it achieves a time complexity that is linear in the number of "moments" when the occurrence of multiple equilibria does not depend on the number of "moments". Based on this procedure, we construct a consistent estimator of the payoff function, an asymptotically uniformly valid and easy-to-implement test for linear hypotheses on the payoff function, and a consistent method to group payoff functions according to the unobserved heterogeneity. Extensive simulations demonstrate the finite sample efficacy of our procedures.

Keywords: Matching-types problem; Minimum-distance characterization; Multistep moment selection procedure; Time complexity.

*Department of Economics, University of Washington; e-mail: fany88@uw.edu

†MOE Key Lab of Econometrics, WISE, Department of Statistics and Data Science at School of Economics, Xiamen University; e-mail: jiangshuo@xmu.edu.cn

‡School of Economics, the University of Sydney; e-mail: xuetao.shi@sydney.edu.au

1 Introduction

Motivation and Main Contributions The sequential approach to identification and estimation of discrete games of incomplete information is widely used in the literature; see Aguirregabiria and Mira (2007), Bajari et al. (2007) and Pesendorfer and Schmidt-Dengler (2008) for seminal contributions and Bajari et al. (2013) for a survey. In the first step, the equilibrium conditional choice probabilities (CCPs hereafter) at each observed state are identified and estimated from the data. In the second step, the payoff function is identified and estimated using variations in the observed state variable such as exclusion restrictions.[1] By avoiding the computation of equilibrium for every given state and parameter value, the sequential approach is computationally less costly than the all-solution or the joint method such as the nested fixed-point algorithm. However, the sequential approach relies critically on the assumption that there is no common knowledge payoff-relevant unobserved heterogeneity (unobserved heterogeneity hereafter) in the payoff function.[2] As discussed extensively in Aguirregabiria and Mira (2019), this assumption is very restrictive and often violated in the data, motivating them to study identification of a general class of games of incomplete information with payoff relevant unobserved heterogeneity and multiple equilibria.

As pointed out in Aguirregabiria and Mira (2019), the presence of payoff relevant unobserved heterogeneity creates a major challenge in the sequential identification referred to as the matching-types problem, i.e., the difficulty of correctly matching equilibrium CCPs for each unobserved state across different observed states. The matching-types problem arises because the first-step identification of equilibrium CCPs at each observed state is equivalent to the identification of a nonparametric finite mixture model, which is known to be identified only up to a label swapping of the mixing components. Despite this matching-types problem, Aguirregabiria and Mira (2019) establish a necessary and sufficient condition for the sequential identification of model primitives in games with unobserved heterogeneity of finite support and multiple equilibria. Building on their sequential identification result, this paper develops a sequential estimation approach for the class of games of incomplete information allowing for both unobserved heterogeneity of finite support and multiple

[1] Depending on the contexts/specifications, we will use payoff function, payoff vector, and payoff parameter interchangeably throughout this paper.

[2] In De Paula (2013), such unobserved heterogeneity is also called game-level heterogeneity or game-level shock.

equilibria.[3]

To tackle the matching-types problem, we construct a novel characterization of the correct matching and the true payoff vector via a minimum-distance criterion with both correct and incorrect "moments". The set of correct moments corresponds to the correct matching; and the true payoff vector is uniquely determined through the correct matching. In the new minimum-distance criterion, the moment functions are linear in the unknown payoff vector with coefficients depending on the equilibrium CCPs identified in the first step. Estimation of the equilibrium CCPs is standard and can be done expeditiously using existing methods such as those in Bonhomme et al. (2016) and Xiao (2018). Using the plug-in estimators of the coefficients, we obtain a vector of moment functions, based on which the correct set of moments or the correct matching will be selected and the payoff vector will be estimated. Although this procedure falls within the general framework of Andrews (1999), the moment selection procedures in Andrews (1999) are computationally costly and oftentimes infeasible even in games with moderately sized state spaces. The reason is that the number of matchings grows exponentially with the size of the state space. For example, in the *Simple Game* introduced in Section 2 with medium numbers of players and observed and latent states, there can be *thousands of trillions* of matchings.

To overcome this computational challenge, we propose a new and ingenious *multistep moment selection* (MMS) procedure for selecting the correct matching. It is based on the insight that in the minimum-distance criterion, a correct matching selects the same latent state across all observed states. As a result, a mismatch on any single observed state results in a wrong matching that needs not to be considered in the estimation. Exploiting this feature, in the new MMS procedure, we first eliminate matchings that are incorrect with high probability in multiple steps, then estimate the correct matching and the payoff vector using the remaining possible matchings. By carefully designing the steps involved, the new MMS procedure selects the correct matching with probability approaching one and is much faster to implement than the moment selection procedures in Andrews (1999) for games with large state spaces. Theoretically, we show that when there is no multiple equilibria or when the number of observed states with multiple equilibria does not increase with

[3]Although the joint identification does not suffer from the matching-types problem, estimators based on it for games with both unobserved heterogeneity and multiple equilibria have not been formally developed and are unexplored in practice. Additionally, they would carry the heavy computational burden of the existing nested fixed-point algorithm.

the number of moments, the new MMS procedure achieves a linear time complexity in the number of moments for large sample sizes.[4] This is a significant improvement over the exponential time complexity of the moment selection procedures in Andrews (1999). Practically, the new MMS-based estimator of the payoff vector can be calculated within *a second* when there are *thousands of trillions* of matchings; while the estimator in Andrews (1999) requires *thousands of seconds* to compute even when there are only *millions* of matchings.

When there are multiple equilibria played in the data, both multiple equilibria and unobserved heterogeneity contribute to the mixtures of CCPs. To estimate the cardinality of the support of the unobserved heterogeneity and the payoff vector on each latent state, we propose a method for grouping the payoff vectors by extending the k-means method to penalize more clusters. The number of clusters estimates the number of latent states, and the payoffs on each latent state are estimated by the centers of each cluster. We show that our estimators consistently estimate the number of latent states and the payoff vector on each latent state. This method of separating multiple equilibria from unobserved heterogeneity is novel and can be used in other contexts such as the dynamic game with both multiple equilibria and unobserved heterogeneity studied in Luo et al. (2022).

Lastly, we develop a fast-to-compute and asymptotically uniformly valid inference procedure for linear hypotheses on the payoff vector. Despite the difficulties in general post-selection inference, our test is asymptotically uniformly valid and is easy to implement with known critical values from the chi-squared distribution.

Although we focus on the class of games with nonparametric payoff functions in Aguirregabiria and Mira (2019), we demonstrate that the novel minimum-distance characterization can be easily modified for games with parametric payoff functions commonly adopted in empirical work. As a result, the MMS procedure applies to these games as well. In the simulation section, we report the finite sample performance of the new MMS procedure, the estimators for the payoffs and the number of unobserved heterogeneity types, and the test when applied to four games with parametric payoffs. Based on the simulation results, we provide a rule-of-thumb for choosing the tuning parameters for implementation of the MMS procedure, and demonstrate its effectiveness using different designs constructed from the four games. Overall, the simulation results confirm the finite sample efficacy of both the estimation and

[4]The time complexity describes the amount of time it takes to run an algorithm. It is commonly estimated by counting the number of elementary operations performed by the algorithm.

inference procedures.

Related Literature Our paper connects with several strands of the literature. Firstly, the paper is closely related to works on static games of incomplete information with/without unobserved heterogeneity. In the analysis of strategic timing incentives among radio stations, Sweeting (2009) estimates a parametric game of incomplete information that allows for multiple equilibria but no unobserved heterogeneity. He also states that "estimation of a game with many possible choices, multiple equilibria, and observed and possibly unobserved heterogeneity is well beyond the current literature." Sweeting (2009) has sparked important research. De Paula and Tang (2012) propose a formal test for multiple equilibria when there is no unobserved heterogeneity. Grieco (2014) studies identification and estimation in a game with normally distributed private information, allowing both for the presence of multiple equilibria and for normally distributed unobserved heterogeneity. In Grieco (2014), the unobserved heterogeneity is of a nuisance nature; and the parameter of interest does not depend on unobserved heterogeneity. Xiao (2018) studies sequential identification and estimation in a game with multiple equilibria and no unobserved heterogeneity. She develops a new method based on eigendecomposition for identifying and estimating CCPs that come from multiple equilibria. Aguirregabiria and Mira (2019) present a general identification framework allowing for both unobserved heterogeneity and multiple equilibria, which nests the setup in Xiao (2018). In Aguirregabiria and Mira (2019), the payoff parameter of interest is allowed to depend on unobserved heterogeneity, which differs from the parameter of interest in Grieco (2014). Luo et al. (2022) show that the matching-types problem in dynamic games of incomplete information could be resolved by making use of the special structure of Markov perfect equilibrium and the longitudinal variations of observed states. There is another strand of literature on the identification and estimation of complete information games such as Bajari et al. (2010), in which all unobservables are common knowledge among players. A recent paper by Magnolfi and Roncoroni (2023) studies identification and estimation under the solution concept of the Bayesian Correlated Equilibrium, which is proposed by Bergemann and Morris (2013, 2016). Magnolfi and Roncoroni (2023) allow for all information structures consistent with players knowing their own payoffs and the distribution of opponents' payoffs. Their information structure nests both

the complete and incomplete information settings.[5]

Our paper also contributes to the literature on moment selection and uniform inference. In a seminal paper, Andrews (1999) proposes several consistent moment selection procedures for generalized method of moments estimation with valid and invalid moments. Andrews and Lu (2001) extend these procedures and apply them to dynamic panel data models. Two problems remain unsolved regarding the moment selection procedures. First, it is well-known that executing the procedures in Andrews (1999) and Andrews and Lu (2001) can be computationally costly (see Liao (2013)).[6] Second, the built-in moment selection implies that the inference problem is a post-selection inference. Constructing an asymptotically uniformly valid inference method is challenging in this context (see Leeb and Pötscher (2005) and Leeb and Pötscher (2008)). In the paper, we solve both problems by fully exploiting the structure of our setup. First, a correct matching selects the same latent state for each observed state. As a result, we know the structure of the set of correct moments. This fact enables us to design a multistep algorithm that is fast to compute. Second, in our setting, the minimum number of correct moments is known. This allows us to construct an asymptotically uniformly valid and easy-to-implement inference method.

Organization of the Rest of this Paper The rest of this paper is organized as follows. Section 2 uses two games of incomplete information with unobserved heterogeneity to introduce our novel minimum-distance criterion of the payoff vector. The first game is a member of the class of games with nonparametric payoff functions studied in Aguirregabiria and Mira (2019) and is referred to as the Simple Game; and the second game is the same as the Simple Game except that its payoff function is parameterized. Section 3 proposes the novel MMS procedure and the estimator of the payoff vector, proves its consistency, and shows the asymptotic linear time complexity of the procedure for the Simple Game. Section 4 develops an asymptotically uniformly valid test for linear hypotheses on the payoff vector. Section 5 extends the methods developed for the Simple Game to a game with a general number of players, actions, latent states, and most importantly multiple equilibria referred to as the General Game. We extend the MMS estimation procedure and the asymptotically uniformly valid test developed for the Simple Game to the General Game. In Section

[5] Haile and Tamer (2003) and Aradillas-López et al. (2016) study identification and inference of auction models under weak assumptions that could allow for multiple equilibria.

[6] Unlike Liao (2013) or Cheng and Liao (2015), a known set of valid moments that guarantee identification of the unknown parameter is not available in our setup.

6, we introduce variants of the Simple Game and General Game and investigate the finite sample performance of our estimation and inference methods via Monte Carlo simulation. Section 7 concludes. Mathematical details of the results for the General Game are provided in Appendix A. Additional materials are collected in the online appendix. Appendix B contains proofs for the results in the paper. Appendix C contains further details on Xiao (2018)'s CCP estimator and the identification in the General Game. Appendix D contains additional details on the simulation. The codes for implementing the multistep estimation and inference procedures are available at https://github.com/FanJiangShi/MMSP.

We close this section by introducing some notations used throughout this paper. For any $q \times 1$ vector E, let $\|E\|$ denote its Euclidean norm and $\|E\|_0$ denote its L_0 norm, i.e., the number of non-zero elements in E. For Ω being some $q \times q$ matrix, denote $\|E\|_\Omega^2 \equiv E^\top \Omega E$. For any finite set, $|\cdot|$ denotes its cardinality. For any given q-dimensional vector c of zeros and ones and some $q \times p$ matrix A, let A_c denote the submatrix of A generated by deleting the rows in A corresponding to zeros in c. Let I_q be a $q \times q$ identity matrix. "wp $\to 1$" denotes "with probability approaching one".

2 A Minimum-Distance Characterization of the Payoff Vector

In this section, we use two games of incomplete information with unobserved heterogeneity to introduce our novel minimum-distance characterization of the payoff vector. The first game is a member of the class of games with nonparametric payoff functions studied in Aguirregabiria and Mira (2019) and is referred to as the Simple Game. The second game is the same as the Simple Game except that its payoff function is parameterized. We introduce the Simple Game in Section 2.1 and review its sequential identification in Section 2.2. In Section 2.3, we construct a novel characterization of the payoff function in the Simple Game via a minimum-distance criterion with both correct and incorrect moments. In Section 2.4, we introduce the payoff function of the second game and show how the minimum-distance characterization accommodates for the parametric structure of the payoff function of the second game.

2.1 The Simple Game

In the Simple Game, there are *three* players, *two* actions, *one* exclusive observed state variable, and *one* dichotomous unobserved state variable.[7] Each player, denoted as $i = 1, 2, 3$, chooses an action $d_i \in \{0, 1\}$. Before choosing his action, player i draws his private information $\widetilde{\epsilon}_i(d_i)$ for two actions $d_i = 0$ and $d_i = 1$ from a bivariate distribution. For $i = 1, 2, 3$, denote $z_i \in \mathscr{Z}_i$ as an observable exclusive state variable which does not enter the payoffs of other players, where $\mathscr{Z}_i$ is a finite set with cardinality $|\mathscr{Z}_i|$. Let $k \in \mathscr{K} \equiv \{A, B\}$ be a common knowledge state variable that is known by all players but unobserved by the econometrician.[8] Player i's payoff from choosing action d_i is given by $\widetilde{\pi}_i(d_i, \mathbf{d}_{-i}, z_i, k, \widetilde{\epsilon}_i(d_i))$, where the vector $\mathbf{d}_{-i}$ denotes the joint actions of all the other players except i.

Following Aguirregabiria and Mira (2019), we assume that player i's payoff is additively separable in his private information $\widetilde{\epsilon}_i(d_i)$ and can be written as

$$\widetilde{\pi}_i(d_i, \mathbf{d}_{-i}, z_i, k, \widetilde{\epsilon}_i(d_i)) = \widetilde{\pi}_i(d_i, \mathbf{d}_{-i}, z_i, k) - \widetilde{\epsilon}_i(d_i),$$

where $\widetilde{\pi}_i(d_i, \mathbf{d}_{-i}, z_i, k)$ captures how player i's payoff for choosing d_i changes with respect to his opponents' actions and state variables.[9] Since the optimal action is invariant under monotonically increasing transformations of payoffs, we normalize the payoffs using $\widetilde{\pi}_i(0, \mathbf{d}_{-i}, z_i, k, \widetilde{\epsilon}_i(0))$ and define the normalized payoff for $d_i = 1$ as

$$\pi_i(1, \mathbf{d}_{-i}, z_i, k) \equiv \frac{\widetilde{\pi}_i(1, \mathbf{d}_{-i}, z_i, k) - \widetilde{\pi}_i(0, \mathbf{d}_{-i}, z_i, k)}{sd},$$

where sd denotes the standard deviation of $\widetilde{\epsilon}_i(1) - \widetilde{\epsilon}_i(0)$. We refer to $\pi_i(d_i, \mathbf{d}_{-i}, z_i, k)$ as the *payoff function* for player i hereafter. By normalization, $\pi_i(0, \mathbf{d}_{-i}, z_i, k) = 0$.

Define the normalized private information for player i as $\epsilon_i \equiv \frac{1}{sd}(\widetilde{\epsilon}_i(1) - \widetilde{\epsilon}_i(0))$. Let $\mathbf{z} \equiv (z_1, z_2, z_3) \in \mathscr{Z} \equiv \mathscr{Z}_1 \times \mathscr{Z}_2 \times \mathscr{Z}_3$. We adopt the assumption in Aguirregabiria and Mira (2019) on ϵ_i stated below.[10]

[7] It follows from Allman et al. (2009) and Aguirregabiria and Mira (2019) that when the number of mixing components is 2, the minimum number of players required for identifying CCPs up to a label swapping is 3.

[8] The assumption of a fixed and finite support for unobserved heterogeneity is not only used in Aguirregabiria and Mira (2019) (p. 1663), but also in single agent dynamic discrete choice models such as Kasahara and Shimotsu (2009).

[9] The additive separability of the private information is commonly assumed in the literature on the econometrics of games of incomplete information. Examples include Sweeting (2009), De Paula and Tang (2012), Bajari et al. (2013), and Grieco (2014).

[10] Other papers that maintain such an assumption on private information include Zhu and Singh

Assumption 2.1. *(i)* $\{\epsilon_i\}_{i=1}^3 \overset{i.i.d}{\sim} F(\cdot)$*, where* $F(\cdot)$ *is an absolutely continuous distribution function with a probability density function denoted as* $f(\cdot)$ *and is known to the econometrician. (ii) The support of* $f(\cdot)$ *is* $\mathbb{R}$*. (iii)* ϵ_1, ϵ_2*, and* ϵ_3 *are independent of the state variables* $(\mathbf{z}, k)$*.*

A (pure) strategy in this game is defined as follows:

Definition 2.1 (Strategy)**.** *For given* $\mathbf{z}$ *and* k*, a (pure) strategy for player* i *is a mapping* $\sigma_i(\epsilon_i, \mathbf{z}, k) : \mathbb{R} \times \mathscr{Z} \times \mathscr{K} \to \{1, 0\}$*.*

For notational compactness, we use $\boldsymbol{\sigma} \equiv (\sigma_1(\epsilon_1, \mathbf{z}, k), \sigma_2(\epsilon_2, \mathbf{z}, k), \sigma_3(\epsilon_3, \mathbf{z}, k))$ to denote a strategy profile given $(\mathbf{z}, k)$. Let $\mathbb{1}(\cdot)$ denote the indicator function. Any given $\boldsymbol{\sigma}$ is completely characterized by the following CCPs:

$$p_i \equiv \int \mathbb{1}(\sigma_i(\epsilon_i, \mathbf{z}, k) = 1) f(\epsilon_i) d\epsilon_i, \text{ for } i = 1, 2, 3.$$

Denote j and q as the two players other than player i. The *expected payoff function* for player i with $d_i = 1$ for given $(\mathbf{z}, k)$ and $\boldsymbol{\sigma}$ is computed as:[11]

$$\overline{\pi}_i(1, \mathbf{z}, k, \boldsymbol{\sigma}) = \sum_{d_j, d_q \in \{0,1\}} p_j^{d_j} (1 - p_j)^{1-d_j} p_q^{d_q} (1 - p_q)^{1-d_q} \pi_i(1, (d_j, d_q), z_i, k). \quad (2.1)$$

Bayesian Nash Equilibrium (BNE) is then defined as follows:

Definition 2.2 (Equilibrium)**.** *For any given* $(\mathbf{z}, k)$*, a BNE of the game is a strategy profile* $\boldsymbol{\sigma}^*$ *such that for any player* i *and for any* ϵ_i*,*

$$\sigma_i^*(\epsilon_i, \mathbf{z}, k) = \arg \max_{d_i \in \{0,1\}} \{\overline{\pi}_i(d_i, \mathbf{z}, k, \boldsymbol{\sigma}^*) - \epsilon_i\}.$$

In the Simple Game, we assume that the data are rationalized by a single equilibrium. This assumption will be discarded in Section 5.

Assumption 2.2. *A single equilibrium is played in the data for each* $(\mathbf{z}, k) \in \mathscr{Z} \times \mathscr{K}$*.*

(2009), Li et al. (2013), and Xiao (2018). Papers that allow for unknown distribution for private information include Aradillas-Lopez (2010) and Lewbel and Tang (2015). Papers that allow for correlated private information among players include Wan and Xu (2014), Xu (2014), and Liu et al. (2017).

[11] Because of the normalization, $\overline{\pi}_i(0, \mathbf{z}, k, \boldsymbol{\sigma}) = 0$.

Denote the *equilibrium* CCP of choosing action 1 for player i as $p_i(\mathbf{z}, k) \equiv \Pr(d_i = 1 \mid \mathbf{z}, k)$. Then it holds that

$$p_i(\mathbf{z}, k) = \int \mathbb{1}(\sigma_i^*(\epsilon_i, \mathbf{z}, k) = 1) f(\epsilon_i) d\epsilon_i.$$

For any $(\mathbf{z}, k)$, the BNE of the game is equivalently characterized by the equilibrium CCP vector

$$\mathbf{p}(\mathbf{z}, k) \equiv (p_1(\mathbf{z}, k), p_2(\mathbf{z}, k), p_3(\mathbf{z}, k)), \tag{2.2}$$

where

$$p_i(\mathbf{z}, k) = F(\overline{\pi}_i(1, \mathbf{z}, k, \boldsymbol{\sigma}^*)) \text{ for } i = 1, 2, 3.$$

Under Assumption 2.2, the equilibrium expected payoff function is unique and only depends on $\mathbf{z}$ and k. For brevity, we denote it as $\overline{\pi}_i(1, \mathbf{z}, k) \equiv \overline{\pi}_i(1, \mathbf{z}, k, \boldsymbol{\sigma}^*)$.

Remark 2.1. For any given $(\mathbf{z}, k)$ and strategy profile $\boldsymbol{\sigma}$ with corresponding CCP vector (p_1, p_2, p_3), the probability that choice 1 is optimal for player i in the Simple Game is given by $F(\overline{\pi}_i(1, \mathbf{z}, k, \boldsymbol{\sigma}))$ for $i = 1, 2, 3$. This defines the best response mapping for players $i = 1, 2, 3$ on $(\mathbf{z}, k)$ as follows:

$$\Lambda_{i\mathbf{z}k}(\overline{\pi}_i(1, \mathbf{z}, k, \boldsymbol{\sigma})) = F(\overline{\pi}_i(1, \mathbf{z}, k, \boldsymbol{\sigma})). \tag{2.3}$$

In the Simple Game, the best response mapping is simply the cumulative distribution function. Given the values of payoff functions, by (2.1), these best response mappings only depend on CCPs of other players. Aguirregabiria and Mira (2019) characterize the BNE as a fixed point of the best response mapping of all players.

The game together with the equilibrium constitute the underlying structure that generates the data. Following Aguirregabiria and Mira (2019), the form of data available to the econometrician is specified in the following assumption.

Assumption 2.3. *The econometrician observes a random sample on players' actions and observable state variables* $\{(d_{1m}, d_{2m}, d_{3m}, z_{1m}, z_{2m}, z_{3m})\}_{m=1}^n$.

Let $p^A(\mathbf{z}) \equiv \Pr(k = A \mid \mathbf{z})$ and $p^B(\mathbf{z}) \equiv \Pr(k = B \mid \mathbf{z})$. We impose the following assumption on the distribution of the latent state.

Assumption 2.4. *(i) For any* $\mathbf{z} \in \mathscr{Z}$, $0 < p^A(\mathbf{z}) < 1$. *(ii)* $|\mathscr{K}|$ *is known to the econometrician.*

In the Simple Game, $|\mathscr{K}| = 2$. Assumption 2.4 (ii) is made to simplify the exposition. Section 5 allows for unknown $|\mathscr{K}|$.

2.2 A Review of Sequential Identification of the Payoff Vector in the Simple Game

Stacking the payoff function evaluated at all combinations of opponents' actions, we obtain the following *payoff vector* for player i at observed state z_i and unobserved state k:

$$\pi_{iz_ik} \equiv \begin{bmatrix} \pi_i(1,(1,1),z_i,k) \\ \pi_i(1,(0,1),z_i,k) \\ \pi_i(1,(1,0),z_i,k) \\ \pi_i(1,(0,0),z_i,k) \end{bmatrix}. \tag{2.4}$$

The objective is to identify, estimate, and conduct inference on the payoff vectors π_{iz_iA} and π_{iz_iB} for each player i and observed state z_i. Since the identification procedure is exactly the same for all i and z_i, in the following discussion we focus on the payoff vectors for player 1 with $z_1 = z_1^1$, where z_1^1 is the first element in $\mathscr{Z}_1$.

2.2.1 Step-1: Identification of the Equilibrium CCP Vector (up to a Label Swapping)

The identification of the equilibrium CCP vector, $\mathbf{p}(\mathbf{z},k)$ in Equation (2.2), and mixing weights, $p^A(\mathbf{z})$ and $p^B(\mathbf{z})$, makes use of the following system of equations: for $(d_1,d_2,d_3) \in \{0,1\}\times\{0,1\}\times\{0,1\}$ and $\mathbf{z} \in \mathscr{Z}$,

$$p(d_1,d_2,d_3 \mid \mathbf{z}) = \sum_{k\in\{A,B\}} \left[p^k(\mathbf{z}) \prod_{i=1}^{3} (p_i(\mathbf{z},k))^{d_i} (1-p_i(\mathbf{z},k))^{1-d_i} \right], \tag{2.5}$$

where for any (d_1,d_2,d_3), $p(d_1,d_2,d_3 \mid \mathbf{z})$ denotes the conditional probability of players' joint actions identified from the sample information. It is written as a weighted sum of the product of individual CCPs on each unobserved state. Equation (2.5) is a nonparametric finite mixture model. Sufficient conditions for identifiability of such models can be found in Theorem 4 and Corollary 5 in Allman et al. (2009). In the Simple Game, the following assumption guarantees identification in the first step.

Assumption 2.5. $\mathbf{P}_{i\mathbf{z}}$ *has full rank for any* $\mathbf{z} \in \mathscr{Z}$ *and* $i = 1,2,3$, *where*

$$\mathbf{P}_{i\mathbf{z}} \equiv \begin{bmatrix} p_i(\mathbf{z},k) & p_i(\mathbf{z},k') \\ 1-p_i(\mathbf{z},k) & 1-p_i(\mathbf{z},k') \end{bmatrix} \text{ for } k,k' \in \{A,B\} \text{ and } k \neq k'.$$

Assumption 2.5 is the same as Assumption (d) in Proposition 1 of Aguirregabiria and Mira (2019) or Condition (2) in Lemma 1 of Xiao (2018). It holds if and only if

$p_i(\mathbf{z}, k) \neq p_i(\mathbf{z}, k')$ for any $\mathbf{z} \in \mathscr{Z}$ and $i = 1, 2, 3$. Constructive identification results for nonparametric finite mixture models have been established by Bonhomme et al. (2016) and Xiao (2018), among others. The approach in Bonhomme et al. (2016) is based on the simultaneous diagonalization of a set of matrices in the same non-orthogonal basis and is applicable to a wide range of multivariate latent-structure models including the finite mixture model and the hidden Markov model. The procedure in Xiao (2018) is based on eigendecomposition of a set of matrices. It is developed for identifying CCPs of discrete games of incomplete information when CCPs could be expressed as a finite mixture model, which is the case in our problem as shown in Equation (2.5).

For each $\mathbf{z} \in \mathscr{Z}$, Step-1 identifies two equilibrium CCP vectors denoted as $\mathbf{p}(\mathbf{z}, k)$ and $\mathbf{p}(\mathbf{z}, k')$, where $k, k' \in \{A, B\}$ and $k \neq k'$. Since k and k' in the CCP vectors $\mathbf{p}(\mathbf{z}, k)$ and $\mathbf{p}(\mathbf{z}, k')$ are unknown, we say that the equilibrium CCP vectors are identified up to a label swapping, i.e., swapping of labels A and B.

2.2.2 Step-2: Identification of the Payoff Vector—the Matching-types Problem

It follows from Section 2.1 that the equilibrium expected payoff function for player 1 given $(\mathbf{z}, k)$ is

$$\overline{\pi}_1(1, \mathbf{z}, k) = F^{-1}(p_1(\mathbf{z}, k)). \tag{2.6}$$

In the rest of this section, we treat the expected payoff function as known (up to a label swapping) since $F(\cdot)$ is known and $p_1(\mathbf{z}, k)$ is identified (up to a label swapping) from Step-1.

For any $\mathbf{z} \in \mathscr{Z}$ and $k \in \{A, B\}$, let $\mathbf{p}_{-1}(\mathbf{z}, k)$ denote the following row vector:

$$[p_2(\mathbf{z}, k) p_3(\mathbf{z}, k), (1 - p_2(\mathbf{z}, k)) p_3(\mathbf{z}, k), p_2(\mathbf{z}, k)(1 - p_3(\mathbf{z}, k)), (1 - p_2(\mathbf{z}, k))(1 - p_3(\mathbf{z}, k))].$$

It consists of joint probabilities of player 1's opponents' actions on latent state k. Let $\mathscr{Z}^1 \equiv \{(z_1^1, z_2, z_3) : z_2 \in \mathscr{Z}_2, z_3 \in \mathscr{Z}_3\}$ be the subset of $\mathscr{Z}$ when holding $z_1 = z_1^1$. To simplify the discussion, denote the cardinality of $\mathscr{Z}^1$ as $l \equiv |\mathscr{Z}_2| \times |\mathscr{Z}_3|$, and let $\mathbf{z}^1, \ldots, \mathbf{z}^l$ be the l elements in $\mathscr{Z}^1$.

Denote the true payoff vectors as $\pi_{0k} \equiv \pi_{1z_1^1 k}$ for $k = A, B$, where $\pi_{1z_1^1 k} \in \Pi$ is defined in (2.4) and $\Pi \subseteq \mathbb{R}^{l_\pi}$ with $l_\pi = 4$ denotes the parameter space of π_{0k}. For $\mathbf{z} \in \{\mathbf{z}^1, \ldots, \mathbf{z}^l\}$, player 1's equilibrium expected payoff can be written as:

$$\overline{\pi}_1(1, \mathbf{z}, k) = \mathbf{p}_{-1}(\mathbf{z}, k) \pi_{0k} \text{ for } k = A, B.$$

As $\mathbf{z}$ varies in $\left\{\mathbf{z}^1, \ldots, \mathbf{z}^l\right\}$, z_1 is held constant at $z_1 = z_1^1$. To identify π_{0k}, we need to also hold k constant at A and B respectively. This would be trivial if A and B were observable. Collecting equations corresponding to A and B separately delivers two systems of equations:

$$\overline{\pi}_1(1, \mathbf{z}, A) = \mathbf{p}_{-1}(\mathbf{z}, A)\, \pi_A \text{ for all } \mathbf{z} \in \left\{\mathbf{z}^1, \ldots, \mathbf{z}^l\right\} \text{ and} \tag{2.7}$$

$$\overline{\pi}_1(1, \mathbf{z}, B) = \mathbf{p}_{-1}(\mathbf{z}, B)\, \pi_B \text{ for all } \mathbf{z} \in \left\{\mathbf{z}^1, \ldots, \mathbf{z}^l\right\}, \tag{2.8}$$

where $\pi_k \in \Pi$ for $k = A, B$. As long as $\left[\mathbf{p}_{-1}(\mathbf{z}^1, A)^\top, \ldots, \mathbf{p}_{-1}(\mathbf{z}^l, A)^\top\right]^\top$ and $\left[\mathbf{p}_{-1}(\mathbf{z}^1, B)^\top, \ldots, \mathbf{p}_{-1}(\mathbf{z}^l, B)^\top\right]^\top$ have full column ranks, systems (2.7) and (2.8) both have unique solutions corresponding to π_{0A} and π_{0B}. However, neither A nor B is observable. Systems like (2.7) and (2.8) are not available after Step-1 because the equilibrium CCP vectors are only identified up to a label swapping. The researcher needs to track the same latent state across different observed states, i.e., solve the matching-types problem.[12]

2.3 A Minimum-Distance Criterion for the Correct Matching and the Payoff Vector for the Simple Game

Proposition 3 in Aguirregabiria and Mira (2019) presents a necessary and sufficient condition for Step-2 identification of the payoff vector: *a correct matching* leads to a unique solution to the system for expected payoffs, and this solution identifies the payoff vector; while *an incorrect matching* delivers no solution. Building on this result, we construct a novel minimum-distance criterion with correct and incorrect moments to characterize the correct matching and the true payoff vector associated with each latent state.

We first stack the expected payoff functions evaluated at two latent states for $\mathbf{z}^1$, and then repeat the step for the all $\mathbf{z}^2, \ldots, \mathbf{z}^l$ to obtain the vector $\overline{\pi}$ of expected payoffs of dimension $2l$. In the same fashion, we construct the coefficient matrix Γ of dimension $2l \times l_\pi$. They are

$$\begin{aligned}
&\overline{\pi} \equiv \left[\overline{\pi}_1\left(1, \mathbf{z}^1, k_1\right), \overline{\pi}_1\left(1, \mathbf{z}^1, k_1'\right), \ldots, \overline{\pi}_1\left(1, \mathbf{z}^l, k_l\right), \overline{\pi}_1\left(1, \mathbf{z}^l, k_l'\right)\right]^\top \text{ and} \\
&\Gamma \equiv \left[\mathbf{p}_{-1}\left(\mathbf{z}^1, k_1\right)^\top, \mathbf{p}_{-1}\left(\mathbf{z}^1, k_1'\right)^\top, \ldots, \mathbf{p}_{-1}\left(\mathbf{z}^l, k_l\right)^\top, \mathbf{p}_{-1}\left(\mathbf{z}^l, k_l'\right)^\top\right]^\top,
\end{aligned} \tag{2.9}$$

[12] As noted by Aguirregabiria and Mira (2019), the matching-types problem also exists in the sequential identification of single agent discrete choice model.

where k_t and k'_t are used to denote the pair of latent states on the t-th observed state for $t = 1, ..., l$. For any $t_1 \neq t_2$, k_{t_1} and k_{t_2} (k'_{t_1} and k'_{t_2}) do not necessarily correspond to the same latent state. After we have identified the equilibrium CCP vectors for each observed state up to a label swapping in Step-1, we can construct $\overline{\pi}$ and Γ, and obtain the following system of $2l$ moment functions in $\pi \in \Pi \subseteq \mathbb{R}^{l_\pi}$:

$$G(\pi) \equiv \overline{\pi} - \Gamma\pi \in \mathbb{R}^{2l}. \tag{2.10}$$

There are l different observed states when holding $z_1 = z_1^1$. Each observed state has two groups of identified CCPs. However, we do not know which group corresponds to latent state A and which group corresponds to latent state B. The aim is to select all components of $G(\cdot)$ that correspond to the same latent state, A or B. We call such a selection a correct matching and use it to recover the payoff vector π_{0k} for $k = A, B$.

Definition 2.3 (Correct matching)**.** *A matching for a latent state is said to be correct if and only if it selects all the components of $G(\cdot)$ with the same underlying latent state; otherwise it is incorrect.*

There are two correct matchings associated with the two latent states, A and B. For each given latent state, we aim at identifying its correct matching. We adopt the moment selection framework of Andrews (1999) to associate each matching with a selection vector, where the true selection vector corresponds to the correct matching.

Definition 2.4 (Selection vector)**.** *A selection vector $c \in \mathbb{R}^{2l}$ consists of l zeros and l ones: a "one" indicates that the corresponding element of $G(\cdot)$ is selected and a "zero" indicates that the corresponding element of $G(\cdot)$ is not selected.*

Let $c_0 \in \mathscr{C}$ denote the true selection vector for a given latent state, where $\mathscr{C}$ is the parameter space of c_0 defined below. We identify the correct matching via c_0. Since there are two correct matchings, we define two parameter spaces corresponding to the two latent states. To identify the true selection vectors for both latent states, we let all selection vectors in the first parameter space select the first latent state on the first observed state, and all selection vectors in the second parameter space select the second latent state on the first observed state. Given a specified latent state (first or second) for the first observed state $\mathbf{z}^1$, we select one latent state from each subsequent observed state, $\mathbf{z}^2, \ldots, \mathbf{z}^l$, to match it, leading to the following definition of the parameter space for the true selection vector:

Definition 2.5 (Parameter space of the true selection vector)**.** *The parameter spaces of the true selection vectors for the two latent states are defined as*

$$\mathscr{C}^1 \equiv \left\{ [c_1, \ldots, c_l]^\top \in \mathbb{R}^{2l} : c_1 = [1,0] \text{ and } c_t \in \{[1,0],[0,1]\} \text{ for } t \in \{2,...,l\} \right\} \text{ and}$$
$$\mathscr{C}^2 \equiv \left\{ [c_1, \ldots, c_l]^\top \in \mathbb{R}^{2l} : c_1 = [0,1] \text{ and } c_t \in \{[1,0],[0,1]\} \text{ for } t \in \{2,...,l\} \right\}.$$

Based on Definition 2.5, both $\mathscr{C}^1$ and $\mathscr{C}^2$ have 2^{l-1} elements. Given $\mathscr{C}^1$ or $\mathscr{C}^2$, we call the l moments in $G(\cdot)$ selected by the correct matching the correct moments and the remaining l moments in $G(\cdot)$ incorrect moments. By definition, an incorrect matching selects at least one incorrect moments. To simplify our analysis below, we focus on $\mathscr{C}^1$ and remove the superscript 1.

Denote the selected moment functions from $G(\cdot)$ by a selection vector $c \in \mathscr{C}$ as $G_c(\cdot)$:

$$G_c(\pi) = \overline{\pi}_c - \Gamma_c \pi,$$

where $\overline{\pi}_c \in \mathbb{R}^l$, $\Gamma_c \in \mathbb{R}^{l \times l_\pi}$, and $G_c(\cdot)$ is of dimension l. Let $\pi_0 \in \Pi$ denote the true payoff vector that correspond to the first latent state on the first observed state. Conditions for the identification of (c_0, π_0) are specified as the following.

Assumption 2.6. *(i)* Γ_{c_0} *has full column rank. (ii)* $\text{rank}([\overline{\pi}_c, \Gamma_c]) > \text{rank}(\Gamma_c)$ *for any* $c \neq c_0$*, where* $[\overline{\pi}_c, \Gamma_c]$ *denotes the matrix that combines* $\overline{\pi}_c$ *and* Γ_c*.*

Assumption 2.6 is sufficient for identifying (c_0, π_0), but not necessary. It is stronger than the necessary and sufficient condition discussed in Proposition 3 in Aguirregabiria and Mira (2019) because we consider identification for each pair of player and exclusive state separately for illustrative convenience. The estimation and inference procedures based on the minimum-distance criterion to be introduced in the next sections can be directly applied if the condition in Proposition 3 of Aguirregabiria and Mira (2019) is imposed. The system would stack moment functions for three players together; and the correct matching would select the same latent state on all observed states for all players.

A necessary condition for Assumption 2.6 is that $l \geq l_\pi$. Practically, this requires that for each player, the number of values for the excluded variables of the other players be at least as large as the number of different action profiles for the other players. Since $\overline{\pi}_{c_0} - \Gamma_{c_0}\pi = \mathbf{0}$ has at least one solution, which is the true payoff vector, the assumption on Γ_{c_0} having full column rank is sufficient for $\overline{\pi}_{c_0} - \Gamma_{c_0}\pi$ to have a unique solution. For any $c \neq c_0$, the system of linear equations: $\overline{\pi}_c - \Gamma_c \pi = \mathbf{0}$ has no

solution if and only if rank $([\overline{\pi}_c, \Gamma_c]) >$ rank (Γ_c). Therefore, any system based on an incorrect matching has no solution under Assumption 2.6 (ii). Like the necessary and sufficient condition in Aguirregabiria and Mira (2019), Assumption 2.6 is a high-level condition. In Section 6.1, we numerically illustrate that the assumption holds for all payoffs except for a set of Lebesgue measure zero when $l > l_\pi$.

Lemma 2.1. *Suppose Assumptions 2.1-2.6 hold. Then (i) the system $G_{c_0}(\pi) = 0$ has a unique solution π_0; (ii) for any $c \neq c_0$, the system $G_c(\pi) = 0$ has no solution; and (iii) (c_0, π_0) is identified as the unique minimizer of* $\min_{c \in \mathscr{C}, \pi \in \Pi} \|G_c(\pi)\|^2$.

We use an example to illustrate Lemma 2.1. Suppose the first latent state on $\mathbf{z}^1$ is latent state A. Then c_0 selects components of $\overline{\pi}$ and Γ defined in (2.9) such that $\overline{\pi}_{c_0} = \left[\overline{\pi}_1(1, \mathbf{z}^1, A), \ldots, \overline{\pi}_1(1, \mathbf{z}^l, A)\right]^\top$ and $\Gamma_{c_0} = \left[\mathbf{p}_{-1}(\mathbf{z}^1, A)^\top, \ldots, \mathbf{p}_{-1}(\mathbf{z}^l, A)^\top\right]^\top$. Part (i) requires that $G_{c_0}(\pi) \equiv \overline{\pi}_{c_0} - \Gamma_{c_0}\pi = 0$ has a unique solution, which identifies π_{0A}. Part (ii) of the lemma is about selection vector $c \neq c_0$. An example of such c could select latent state A except for the last observed state, i.e., $\overline{\pi}_c = \left[\overline{\pi}_1(1, \mathbf{z}^1, A), \ldots, \overline{\pi}_1(1, \mathbf{z}^l, B)\right]^\top$ and $\Gamma_c = \left[\mathbf{p}_{-1}(\mathbf{z}^1, A)^\top, \ldots, \mathbf{p}_{-1}(\mathbf{z}^l, B)^\top\right]^\top$. Part (ii) requires that $G_c(\pi) \equiv \overline{\pi}_c - \Gamma_c\pi = 0$ has no solution.

Note that Lemma 2.1 (i) and (ii) imply that the true payoff vector is only defined for the system under the correct matching. In consequence, the correct matching and the true payoff vector can be characterized via a minimum-distance criterion as in Lemma 2.1 (iii). This allows us to address the matching-types problem by selecting the correct moment functions from $G(\cdot)$ and estimating the true payoff jointly.

2.4 Variants of the Simple Game with Parametric Payoff Functions

In empirical work, payoff functions are often parameterized. Since the identification of the equilibrium CCP vector (up to a label swapping) in Step-1 does not depend on the payoff function of the game, sequential identification of parameters in a parameterized payoff function suffers from the same matching-types problem as the Simple Game. In this section, we demonstrate via a variant of the Simple Game that with appropriately redefined $\overline{\pi}$, Γ, and π in (2.10), Lemma 2.1 holds for games with parametric payoffs. Additional variants will be introduced in Section 6.

Consider the following payoff function for player i when choosing $d_i = 1$ is param-

eterized as:

$$\pi_i\left(1, \mathbf{d}_{-i}, z_i, k\right) = z_i\theta_{ik} + \delta_{ik}\left(\sum_{j\neq i} d_j\right), \ k = A, B,$$

where $(\theta_{ik}, \delta_{ik})$ are the parameters of interest. In particular, θ_{ik} measures the effect of the exclusive state variable z_i; and δ_{ik} measures the strategic effect. Both effects are affected by the realization of the latent variable k.

Assume that the distribution of the latent variable conditioning on the observed state is

$$p^A\left(\mathbf{z}\right) = \begin{cases} \frac{3}{4} - \frac{1}{10\sum_{i=1}^{3}|z_i|} & \text{for } \mathbf{z} \neq \mathbf{z}^\dagger \\ \frac{1}{2} & \text{for } \mathbf{z} = \mathbf{z}^\dagger \end{cases},$$

for some $\mathbf{z}^\dagger$. After Step-1 identification of equilibrium CCPs, this game suffers from the matching-types problem in Step-2 because when $\mathbf{z} = \mathbf{z}^\dagger$, components A and B have equal weights and they cannot be separated using their weights.

In contrast to the Simple Game, we can pool all the observed states together to increase the identification power and obtain one system for each player because of the parametric payoff function. Let $l = |\mathscr{Z}_1| \times |\mathscr{Z}_2| \times |\mathscr{Z}_3|$. For example, the system for player 1 is given by

$$G\left(\theta_1, \delta_1\right) = \begin{bmatrix} \overline{\pi}\left(\mathbf{z}^1\right) \\ \vdots \\ \overline{\pi}\left(\mathbf{z}^l\right) \end{bmatrix} - \begin{bmatrix} \Gamma\left(\mathbf{z}^1\right) \\ \vdots \\ \Gamma\left(\mathbf{z}^l\right) \end{bmatrix} \begin{bmatrix} \theta_1 \\ \delta_1 \end{bmatrix},$$

where

$$\overline{\pi}\left(\mathbf{z}^t\right) = \begin{bmatrix} \overline{\pi}_1\left(1, \mathbf{z}^t, k_t\right) \\ \overline{\pi}_1\left(1, \mathbf{z}^t, k'_t\right) \end{bmatrix} \text{ and } \Gamma\left(\mathbf{z}^t\right) = \begin{bmatrix} z_1 & p_2\left(\mathbf{z}^t, k_t\right) + p_3\left(\mathbf{z}^t, k_t\right) \\ z_1 & p_2\left(\mathbf{z}^t, k'_t\right) + p_3\left(\mathbf{z}^t, k'_t\right) \end{bmatrix}.$$

The minimum-distance characterization of the payoff vector we constructed for the Simple Game in Section 2.3 is valid with

$$\overline{\pi} = \begin{bmatrix} \overline{\pi}\left(\mathbf{z}^1\right) \\ \vdots \\ \overline{\pi}\left(\mathbf{z}^l\right) \end{bmatrix}, \ \Gamma = \begin{bmatrix} \Gamma\left(\mathbf{z}^1\right) \\ \vdots \\ \Gamma\left(\mathbf{z}^l\right) \end{bmatrix}, \text{ and } \pi = \begin{bmatrix} \theta_1 \\ \delta_1 \end{bmatrix},$$

where the vector $\overline{\pi}$ is of dimension $2l \times 1$, Γ is the coefficient matrix of dimension $2l \times 2$, and π is the vector with dimension 2×1. Given some selected latent state for $\mathbf{z}^1$, we could match this latent state across all observed states under the same conditions to those for Step-2 identification of the Simple Game in Section 2.3.

3 Multistep Moment Selection Estimation of the Payoff Vector in the Simple Game

As we demonstrate in Section 2.4, the parameters of interest in the variants of the Simple Game share the same minimum-distance characterization as the Simple Game with redefined $\overline{\pi}$, Γ, and π in the moment function. Without loss of generality, we focus on developing estimation and inference procedures for the Simple Game in Sections 3 and 4.

From Lemma 2.1, it follows that the true selection vector c_0 and the true parameter vector π_0 satisfy

$$(c_0, \pi_0) = \arg \min_{c \in \mathscr{C}, \pi \in \Pi} \|G_c(\pi)\|^2_{W(c)}, \tag{3.1}$$

where $G_c(\pi) = \overline{\pi}_c - \Gamma_c \pi$ and $W(c)$ is a positive definite weighting matrix that can depend on c. The matrices $\overline{\pi}$ and Γ depend on equilibrium CCPs and can be estimated by the plug-in approach using Equations (2.6) and (2.9) once the equilibrium CCPs on all the observed and latent states are estimated. Existing methods such as those in Bonhomme et al. (2016) and Xiao (2018) can be used to estimate the equilibrium CCPs. We focus on the CCP estimator developed by Xiao (2018) in this paper and present the steps in Appendix C.1. Denote the resulting estimators as $\overline{\pi}_n$ and Γ_n. We note that this step is standard in the literature and can be implemented fast even for large l. This is because the eigen-decomposition procedure is done for each observed state separately; and each procedure is fast to compute.[13] We focus on the second step from now on.

Let the sample moment functions be

$$G_n(\pi) \equiv \overline{\pi}_n - \Gamma_n \pi, \text{ for } \pi \in \Pi.$$

We use $G_{n,c}(\pi)$ to denote the sample moment functions selected by c. The sample version of (3.1) is

$$(\widetilde{c}, \widetilde{\pi}) \equiv \arg \min_{c \in \mathscr{C}, \pi \in \Pi} \|G_{n,c}(\pi)\|^2_{W_n(c)}, \tag{3.2}$$

where $W_n(c)$ is the sample weighting matrix. This is equivalent to the moment selection procedures in Andrews (1999) when applied to the Simple Game. Solving (3.2) requires performing discrete optimization over $\mathscr{C}$. Since $|\mathscr{C}| = 2^{l-1}$, implementing $(\widetilde{c}, \widetilde{\pi})$ is computationally challenging for large l. For example, when there are

[13] Using one thousand simulations, the average time needed to compute the eigen-decomposition and obtain player 1's CCP on one observed state is less than 10^{-4} second.

seven exclusive states for each player, $l = |\mathscr{Z}_2| \times |\mathscr{Z}_3| = 49$. The size of $\mathscr{C}$ becomes $2^{48} \approx 2.8 \times 10^{14}$. This motivates our computationally less costly multistep moment selection (MMS) procedure proposed in Section 3.1. In Section 3.2, we show consistency of the MMS procedure and its time complexity result. In contrast to the estimator $(\widetilde{c}, \widetilde{\pi})$, which has an exponential time complexity in l, the time complexity of the MMS is asymptotically linear in l. Section 3.3 provides some guidance on the practical implementation of the MMS procedure. Proofs of the theorems in this section are provided in the online appendix.

3.1 Multistep Moment Selection Procedure

As the number of exclusive states $|\mathscr{Z}_2|$ or $|\mathscr{Z}_3|$ increases, the value of l rises sharply. The MMS procedure explores an important feature of the game to reduce the computation time: the true payoff vector π_0 is only defined for the system selected by c_0, and none of the other systems selected by $c \in \mathscr{C}$ and $c \neq c_0$ has a solution. Specifically, in the MMS procedure, we eliminate the matchings/selection vectors $c \in \mathscr{C}$ that are certainly incorrect in *multiple steps* instead of *one step* as in the computation of $(\widetilde{c}, \widetilde{\pi})$. With careful design of the steps involved, we are able to construct an *effective parameter space* for c_0 of a much smaller size than $\mathscr{C}$ (see the last step in the procedure). The MMS procedure selects the correct matching and estimates the payoff vector using the effective parameter space for c_0. To succinctly introduce our idea, we present a two-step moment selection (TMS) procedure first and then extend it to the general MMS procedure.

For any vector sc of dimension $2l$ consisting of zeros and ones, we use $G_{n,sc}(\pi)$ to denote the moment functions selected by sc from $G_n(\pi)$. Different from the selection vectors in $\mathscr{C}$, we let sc select fewer than l moments and call it a sub-selection vector. We partition sc into l subvectors, where each subvector contains two elements. Denote sc_t for $t = 1, ..., l$ as the t-th subvector of sc such that $sc \equiv [sc_1, \ldots, sc_l]^\top$. Define

$$J_n(sc) \equiv \min_{\pi \in \Pi} \|G_{n,sc}(\pi)\|^2 .$$

3.1.1 The TMS Procedure

Let $l_1 \in \{l_\pi, l_\pi + 1, .., l\}$. Heuristically, if we know that the first l_1 moments selected by some selection vector contains incorrect moments, then all the 2^{l-l_1} selection vectors in $\mathscr{C}$ that select the same first l_1 moments can be ignored in the estimation of (c_0, π_0), because a matching is correct only when all the l moments are selected correctly.

Step 1 below identifies matchings of the first l_1 moments that are incorrect with high probability (wp $\to 1$). In Step 2, we estimate the correct matching and the true payoff vector by minimizing an objective function over the product space of the effective parameter space which excludes the incorrect matchings identified in Step 1 and the parameter space Π. We present the detailed steps below.

Step 0: Set $l_1 \in \{l_\pi, l_\pi + 1, .., l\}$, $\alpha_1 \in (0, 1]$, and $\lambda \in (-1, 0)$.

Step 1: Define the collection of sub-selection vectors in Step 1 as

$$\mathscr{SC}^1 \equiv \left\{ \begin{array}{l} [sc_1, \ldots, sc_l]^\top \in \mathbb{R}^{2l} : sc_1 = [1,0]\,; sc_t \in \{[1,0],[0,1]\}\,, \\ \text{for } t \in \{2, ..., l_1\}\,; \text{ and } sc_t = [0,0] \text{ for } t \in \{l_1 + 1, ..., l\} \end{array} \right\}.$$

By definition, $sc^1 \in \mathscr{SC}^1$ selects none of the last $2(l - l_1)$ moments. Sort $J_n(sc^1)$ for all $sc^1 \in \mathscr{SC}^1$, and denote $J_n^{\alpha_1}$ as the value of the $100\alpha_1\%$ smallest. Compare $J_n^{\alpha_1}$ with n^λ. If $J_n^{\alpha_1} > n^\lambda$, then collect all sc^1 such that $J_n(sc^1) \leq J_n^{\alpha_1}$. Otherwise collect all sc^1 such that $J_n(sc^1) \leq n^\lambda$. Denote the collection as $\mathscr{SC}^1_n$:

$$\mathscr{SC}^1_n \equiv \left\{ sc^1 \in \mathscr{SC}^1 : J_n\left(sc^1\right) \leq \max\left\{ J_n^{\alpha_1}, n^\lambda \right\} \right\}.$$

The set $\mathscr{SC}^1_n$ is the output of Step 1.

Step 2: Define the *effective parameter space* for c_0 as

$$\mathscr{C}_n \equiv \left\{ \begin{array}{l} [c_1, \ldots, c_l]^\top \in \mathbb{R}^{2l} : [c_1, \ldots, c_{l_1}] = \left[sc_1^1, \ldots, sc_{l_1}^1\right] \text{ for some} \\ sc^1 \in \mathscr{SC}^1_n; \text{ and } c_t \in \{[1,0],[0,1]\} \text{ for } t \in \{l_1 + 1, ..., l\} \end{array} \right\}.$$

The TMS estimator is defined by the following minimization problem:[14]

$$(\widehat{c}, \widehat{\pi}) \equiv \arg \min_{c \in \mathscr{C}_n, \pi \in \Pi} \left\| G_{n,c}(\pi) \right\|^2_{W_n(c)}.$$

For each $c \in \mathscr{C}_n$, the first $2l_1$ components of c are the same as the first $2l_1$ components of some $sc^1 \in \mathscr{SC}^1_n$; and the last $2(l - l_1)$ components can select any combination of the last $2(l - l_1)$ moments allowed by $\mathscr{C}$. Since every $c \in \mathscr{C}_n$ selects l moments out of the l pairs, $\mathscr{C}_n \subseteq \mathscr{C}$. In the special case where $\mathscr{SC}^1_n = \mathscr{SC}^1$, we have $\mathscr{C}_n = \mathscr{C}$.

Let $sc_0^1 \in \mathscr{SC}^1$ denote the sub-selection vector whose first $2l_1$ elements are the same as c_0. In Step 1, we determine if a sub-selection vector sc is part of an incorrect matching by comparing $J_n(sc)$ with n^λ, because $J_n(sc_0^1) < n^\lambda$ occurs with high probability when n is large. At the same time, we keep at least $100\alpha_1\%$ elements in $\mathscr{SC}^1$ to prevent sc_0^1 from being eliminated because of finite sample error.

[14]The definition of $(\widehat{c}, \widehat{\pi})$ implicitly assumes that the solution to the minimization problem is unique. This can be shown to hold with probability approaching one.

The size of $\mathscr{SC}^1$ is 2^{l_1-1}, which is much smaller than the size of $\mathscr{C}$ when l_1 is smaller than l. As a result, Step 1 can be implemented very fast. The set $\mathscr{SC}_n^1$ has $\alpha_1^\star 2^{l_1-1}$ elements where $\alpha_1^\star \equiv \left|\mathscr{SC}_n^1\right| / \left|\mathscr{SC}^1\right|$; and the size of the effective parameter space $\mathscr{C}_n$ is $\left(\alpha_1^\star 2^{l_1-1}\right) \times 2^{(l-l_1)} = \alpha_1^\star 2^{l-1}$. When $\alpha_1^\star$ is small and l is moderately large, $(\widehat{c}, \widehat{\pi})$ is computationally much faster than $(\widetilde{c}, \widetilde{\pi})$.

Remark 3.1. Setting $\alpha_1 = 1$ recovers $(\widetilde{c}, \widetilde{\pi})$ in (3.2). And $\widehat{\pi} = \widetilde{\pi}$ whenever $\widehat{c} = \widetilde{c}$.

3.1.2 The MMS Procedure

When l is very large, implementing Step 2 above may still be time consuming, because $\alpha_1^\star 2^{l-1}$ can be large. To further reduce the computational time, we extend the above TMS to MMS with any finite number of steps as needed. For example, in an MMS with three steps, the first step is the same as the first step in TMS. Instead of selecting from all the possible combinations of $(2l_1+1)$-th to $2l$-th moments in the second step, we select from the $(2l_1+1)$-th to $2l_2$-th moments, where $l_2 \equiv l_1 + \Delta$ for some prespecified $\Delta \in \{1, .., l-l_1\}$. In the third step, we select from the $(2l_2+1)$-th to $2l$-th moments. Below, we present the detailed procedure for implementing the MMS procedure with $(S+1)$ steps.

Let $\lceil x \rceil$ denote the smallest integer greater than or equal to x.

Step 0: Set $l_1 \in \{l_\pi, l_\pi + 1, .., l\}$, $\alpha_1 \in (0, 1]$, $\lambda \in (-1, 0)$, and $\Delta \in \{1, .., l - l_1\}$. Let $S = \left\lceil \frac{l-l_1}{\Delta} \right\rceil$ and $\alpha = 2^{-\Delta}$.

Step 1: Apply the same procedure as Step 1 in the TMS procedure. The output is the set $\mathscr{SC}_n^1$.

Steps 2, 3, ... S: For $s = 2, ..., S$, define $l_s \equiv l_{s-1} + \Delta$. The input of Step s is the collection of sub-selection vectors defined as

$$\mathscr{SC}^s \equiv \left\{ \begin{array}{c} [sc_1, \ldots, sc_l]^\top \in \mathbb{R}^{2l} : \left[sc_1, \ldots, sc_{l_{s-1}}\right] = \left[sc_1^{s-1}, \ldots, sc_{l_{s-1}}^{s-1}\right] \text{ for some} \\ sc^{s-1} \in \mathscr{SC}_n^{s-1};\ sc_t \in \{[1,0], [0,1]\} \text{ for } t \in \{l_{s-1}+1, ..., l_s\}; \\ \text{and } sc_t = [0,0] \text{ for } t \in \{l_s+1, ..., l\} \end{array} \right\}.$$

By definition, each $sc^s \in \mathscr{SC}_n^s$ consists of three parts: the first $2l_{s-1}$ components of sc^s are the same as the first $2l_{s-1}$ components of some $sc^{s-1} \in \mathscr{SC}_n^{s-1}$; the $(2l_{s-1}+1)$-th to $2l_s$-th components of sc^s select any combinations allowed by $\mathscr{C}$; and sc^s selects none of the last $2(l-l_s)$ moments. Sort $J_n(sc^s)$ for all $sc^s \in \mathscr{SC}^s$, and denote J_n^α as the value of the $100\alpha\%$ smallest. Construct the output of Step s, $\mathscr{SC}_n^s$, as

$$\mathscr{SC}_n^s \equiv \left\{ sc^s \in \mathscr{SC}^s : J_n(sc^s) \leq \max\left\{J_n^\alpha, n^\lambda\right\}\right\}.$$

Step (S + 1): Define the *effective parameter space* for c_0 as

$$\mathscr{C}_n \equiv \left\{ \begin{array}{l} [c_1, \ldots, c_l]^\top \in \mathbb{R}^{2l} : [c_1, \ldots, c_{l_S}] = \left[sc_1^S, \ldots, sc_{l_S}^S\right] \text{ for some} \\ sc^S \in \mathscr{SC}_n^S; \text{ and } c_t \in \{[1,0],[0,1]\} \text{ for } t \in \{l_S+1, ..., l\} \end{array} \right\}.$$

The MMS estimator is defined by the following minimization problem:

$$(\widehat{c}, \widehat{\pi}) \equiv \arg \min_{c \in \mathscr{C}_n, \pi \in \Pi} \|G_{n,c}(\pi)\|^2_{W_n(c)}. \tag{3.3}$$

For each $c \in \mathscr{C}_n$, the first $2l_S$ components of c are the same as the first $2l_S$ components of some $sc^S \in \mathscr{SC}_n^S$; and the last $2(l - l_S)$ components can select any combination of the last $2(l - l_S)$ moments allowed by $\mathscr{C}$.

Let $\alpha_s^\star \equiv |\mathscr{SC}_n^s| / |\mathscr{SC}^s|$. In Step s for $s = 1, ..., S$, the input set $\mathscr{SC}^s$ has the cardinality $2^{l_s - 1} \prod_{i=1}^{s-1} \alpha_i^\star$, and the output set $\mathscr{SC}_n^s$ has the cardinality $2^{l_s - 1} \prod_{i=1}^{s} \alpha_i^\star$. In Step $(S + 1)$, $|\mathscr{C}_n| = 2^{l-1} \prod_{i=1}^{S} \alpha_i^\star$. For large l, we usually have large S and small $\alpha_s^\star$ for $s = 1, ..., S$. The number of optimizations from Step 1 to Step $(S + 1)$ is much fewer than the number of optimizations required for computing $(\widetilde{c}, \widetilde{\pi})$.

Remark 3.2. The choice of the tuning parameters l_1, α_1, λ, and Δ are independent of l and n. See Section 3.3 for more discussion.

Remark 3.3. By replacing the zeros with ones and the ones with zeros in $\widehat{c}$, we obtain an estimator for the true selection vector in $\mathscr{C}^2$. Based on it, we can estimate the payoff vector. Alternatively, we can apply the above MMS procedure to $\mathscr{C}^2$ to obtain the estimators.

3.2 Asymptotic Properties of the MMS Procedure

The consistency of $(\widehat{c}, \widehat{\pi})$ is proved under the following assumptions.

Assumption 3.1. *The space* Π *is compact.*

Assumption 3.2. *For* $\forall\ c \in \mathscr{C}_n$, $W_n(c) \xrightarrow{p} W(c)$ *for some positive definite matrix* $W(c)$.

Assumption 3.2 imposes a standard assumption on the weighting matrix. Note that the weighting matrix $W_n(c)$ is only used in the last step of the MMS procedure. The following theorem states consistency of the estimators $\widehat{c}$ and $\widehat{\pi}$.

Theorem 3.1. *Under Assumptions 2.1-2.6 and 3.1-3.2, it holds that* $\widehat{c} = c_0$ *wp* $\to 1$ *and* $\widehat{\pi} \xrightarrow{p} \pi_0$ *for any* $l_1 \in \{l_\pi, l_\pi + 1, .., l\}$, $\alpha_1 \in (0, 1]$, $\lambda \in (-1, 0)$, *and* $\Delta \in \{1, .., l - l_1\}$.

Theorem 3.1 shows that the MMS procedure is consistent for any l_1, α_1, λ, and Δ that satisfy the requirements in the theorem. The values of the tuning parameters only affect the finite sample performance of the estimator.

The theorem below shows that asymptotically the time and space complexities of the MMS are linear in l.[15]

Theorem 3.2. *Let Assumptions 2.1-2.6 and 3.1 hold. Then with probability approaching one as* $n \to \infty$, *for all payoffs except for a set of Lebesgue measure zero, both the time and space complexities of the MMS procedure are linear in* l.

Consider the space of payoffs for all three players such that the assumptions in the theorem are satisfied. Theorem 3.2 shows that, except for a subset of Lebesgue measure zero in this space, both the computation time and memory storage required for performing the MMS procedure are linear in l with probability approaching one as $n \to \infty$. In other words, except for certain "exceptional" payoffs, the linear time and space complexities hold with high probability when n is large.

Although Theorem 3.2 is an asymptotic result, the simulation results in Section 6 show that the MMS is extremely fast to compute for all DGPs and sample sizes considered. To guarantee the consistency, the MMS cannot eliminate $c \in \mathscr{C}$ if $\min_{\pi \in \Pi} \|G_{n,c}(\pi)\|^2 < n^\lambda$. However, for a small sample size, there might be some $c \neq c_0$ such that the inequality holds.[16] When n becomes larger, fewer selection vectors would satisfy the inequality, because $\min_{\pi \in \Pi} \|G_{n,c}(\pi)\|^2 < n^\lambda$ holds for $\lambda < 0$ only if $c = c_0$ when $n \to \infty$. As a result, $\mathscr{S}\mathscr{C}^s$ in each step contains fewer elements for larger n. The computation time of the MMS decreases and becomes linear in l in the limit.

3.3 Practical Implementation

We summarize the computation of the MMS estimator $(\widehat{c}, \widehat{\pi})$ in the algorithm below and provide guidance on the choice of tuning parameters l_1, α_1, λ, and Δ in finite

[15] The space complexity measures the amount of memory space required for performing an algorithm. It is another important factor when evaluating the efficiency of an algorithm.

[16] Even if in rare cases where $\min_{\pi \in \Pi} \|G_{n,c}(\pi)\|^2$ is small for all $c \neq c_0$, by setting an aggressive λ, the MMS procedure can still improve upon (3.2) in running time.

samples. We discuss the roles of the tuning parameters based on the ascending order of their relative importance to the computational time. For $s = 1, ..., S$, let sc_0^s denote the sub-selection vector whose first $2l_s$ elements are the same as c_0 and the remaining elements are zeros. Namely, sc_0^s selects the correct first l_s moments.

Algorithm 1: the MMS procedure

Input: l_1, α_1, λ, and Δ

Output: $\widehat{c}$ and $\widehat{\pi}$

$S \coloneqq \left\lceil \frac{l-l_1}{\Delta} \right\rceil$, $\alpha \coloneqq 2^{-\Delta}$

/* Start of Step 1 */

Construct $\mathscr{SC}^1$

Compute $J_n\left(sc^1\right)$ for each $sc^1 \in \mathscr{SC}^1$; Sort $J_n\left(sc^1\right)$ for all $sc^1 \in \mathscr{SC}^1$; Find $J_n^{\alpha_1}$ as the $100\alpha_1\%$ smallest

if $J_n^{\alpha_1} > n^\lambda$ **then**

  $\mathscr{SC}_n^1 \coloneqq \left\{sc^1 \in \mathscr{SC}^1 : J_n\left(sc^1\right) \leq J_n^{\alpha_1}\right\}$

else

  $\mathscr{SC}_n^1 \coloneqq \left\{sc^1 \in \mathscr{SC}^1 : J_n\left(sc^1\right) \leq n^\lambda\right\}$

/* End of Step 1. The output is $\mathscr{SC}_n^1$ */

/* Start of Step 2 */

Construct $\mathscr{SC}^2$ from $\mathscr{SC}_n^1$

Compute $J_n\left(sc^2\right)$ for each $sc^2 \in \mathscr{SC}^2$; Sort $J_n\left(sc^2\right)$ for all $sc^2 \in \mathscr{SC}^2$; Find J_n^{α} as the $100\alpha\%$ smallest

if $J_n^{\alpha} > n^\lambda$ **then**

  $\mathscr{SC}_n^2 \coloneqq \left\{sc^2 \in \mathscr{SC}^2 : J_n\left(sc^2\right) \leq J_n^{\alpha}\right\}$

else

  $\mathscr{SC}_n^2 \coloneqq \left\{sc^2 \in \mathscr{SC}^2 : J_n\left(sc^2\right) \leq n^\lambda\right\}$

/* End of Step 2. The output is $\mathscr{SC}_n^2$ */

..... Continue until Step S. The output of Step S is $\mathscr{SC}_n^S$

/* Start of Step $(S+1)$ */

Construct $\mathscr{C}_n$ from $\mathscr{SC}_n^S$

$(\widehat{c}, \widehat{\pi}) \coloneqq \arg\min_{c \in \mathscr{C}_n, \pi \in \Pi} \left\| G_{n,c}\left(\pi\right) \right\|_{W_n(c)}^2$

We first discuss the role of λ. In each step, n^λ serves as a threshold to identify $sc^s \in \mathscr{SC}^s$ that is surely (wp $\to 1$) different from sc_0^s. The threshold is conservative for larger λ and aggressive for smaller λ. By the property of sc_0^s and $G_{n,sc_0^s}\left(\pi\right)$,

$\min_{\pi\in\Pi}\left\|G_{n,sc_0^s}(\pi)\right\|^2 = O_p(n^{-1})$. Thus, if $\min_{\pi\in\Pi}\|G_{n,sc^s}(\pi)\|^2$ does not converge to zero at rate n^λ for $\lambda > -1$ as $n\to\infty$, then $\Pr(sc^s \neq sc_0^s)\to 1$. Excluding such sc^s from the output of Step s, $\mathscr{SC}_n^s$, reduces the number of elements in the input of Step $(s+1)$, $\mathscr{SC}^{s+1}$, and the inputs of all the next steps. We recommend $\lambda = -0.01$ based on the simulation study.

The parameter α_1 acts as a safety net for keeping sc_0^1 in $\mathscr{SC}_n^1$ in finite samples. It also indirectly prevents sc_0^s from being excluded from $\mathscr{SC}_n^s$ in finite samples for $s = 2, ..., S$. Because $\min_{\pi\in\Pi}\left\|G_{n,sc_0^s}(\pi)\right\|^2$ may not be small due to the finite sample error, enough sub-selection vectors need to be included in $\mathscr{SC}_n^s$ so that sc_0^s is not eliminated in each step. We deliberately set $\alpha = 2^{-\Delta}$, so that the number of elements in the output set does not decrease as the algorithm proceeds. There are at least $\alpha_1 2^{l_1-1}$ elements in $\mathscr{SC}_n^s$ after each step. When α_1 is larger, $\mathscr{SC}_n^s$ tends to have more elements, which increases the chance that $sc_0^s \in \mathscr{SC}_n^s$. Extensive simulation suggests setting $\alpha_1 = 0.5\%$.

The tuning parameter l_1 affects the computational time of $(\widehat{c},\widehat{\pi})$, because the set $\mathscr{SC}_n^1$ directly affects $\mathscr{SC}^2$ in Step 2 and $\mathscr{SC}^s$ in all the following steps. There are $\alpha_1^\star 2^{l_1-1}$ elements in $\mathscr{SC}_n^1$, where $\alpha_1^\star \equiv \left|\mathscr{SC}_n^1\right| / \left|\mathscr{SC}^1\right|$, and the same order of elements in $\mathscr{SC}_n^s$ for $s = 2, ..., S$. The value of $\alpha_1^\star \in [\alpha_1, 1]$ is determined by the percentage of sc^1's in $\mathscr{SC}^1$ such that $\min_{\pi\in\Pi}\|G_{n,sc^1}(\pi)\|^2$ is small. If $J_n^{\alpha_1} > n^\lambda$, then $\alpha_1^\star = \alpha_1$, while if $J_n^{\alpha_1} \leq n^\lambda$, then $100\alpha_1^\star\%$ of sc^1's in $\mathscr{SC}^1$ satisfy that $J_n(sc^1) \leq n^\lambda$. In consequence, when only a small portion of elements in $\mathscr{SC}_n^1$ make $\min_{\pi\in\Pi}\|G_{n,sc^1}(\pi)\|^2$ small, $\alpha_1^\star$ is small. Intuitively, $\min_{\pi\in\Pi}\|G_{n,sc^1}(\pi)\|^2$ tends to be small if no or a few incorrect moments are selected by sc^1. Because the percentage of such sc^1's in $\mathscr{SC}^1$ decreases with l_1, $\alpha_1^\star$ is smaller for larger l_1, and vice versa. For example, the percentage of sub-selection vectors in $\mathscr{SC}^1$ that select only one incorrect moment is $(l_1-1)/2^{l_1-1}$, which decreases dramatically as l_1 increases. At the same time, enough moments relative to l_π shall be included in Step 1. Since the number of elements in $\mathscr{SC}_n^1$ is a product of 2^{l_1-1} and $\alpha_1^\star$, we need to balance the two effects of l_1 to achieve faster running time. Based on extensive simulations, we suggest setting $l_1 = 5l_\pi$.[17]

The value of Δ affects the total number of steps and the number of optimizations in each step. Larger Δ leads to fewer steps, because $S = \left\lceil \frac{l-l_1}{\Delta} \right\rceil$. On the other hand, lowering Δ decreases the computation time of each step. Given the number of elements in the output set of Step $(s-1)$, the input set $\mathscr{SC}^s$ of Step s has 2^Δ times

[17] Since l_π is small, for cases where $l < 5l_\pi$, $(\widetilde{c},\widetilde{\pi})$ can be employed directly.

more elements. Decreasing Δ would then reduce the cardinality of $\mathscr{SC}^s$ and the number of operations in each step. We recommend setting $\Delta = 2$ based on extensive simulations. Once Δ is chosen, the value of α is determined accordingly as $2^{-\Delta}$ so that more moments we add in each step, more aggressive we are in the elimination of incorrect matchings.

In summary, we recommend setting $\lambda = -0.01$, $\alpha_1 = 0.5\%$, $l_1 = 5l_\pi$, and $\Delta = 2$ in the MMS procedure. We call such a choice the rule-of-thumb. See more discussion on the roles of the tuning parameters and the rule-of-thumb in Section 6.1.

4 Inference on the Payoff Vector in the Simple Game

This section develops a test for the following linear hypothesis:

$$H_0 : R\pi_0 = r \text{ against } H_1 : R\pi_0 \neq r, \tag{4.1}$$

where R is of dimension $l_R \times l_\pi$ with $\text{rank}\,(R) = l_R$ and r is of dimension $l_R \times 1$. A simple test statistic would be

$$\min_{c \in \mathscr{C}, R\pi = r} \left\| \sqrt{n} G_{n,c}\left(\pi\right) \right\|^2_{W_n(c)},$$

which is expected to be large if the null is incorrect. However, this test statistic can be computationally challenging when the parameter space $\mathscr{C}$ is large. Similar to the multistep estimator proposed in Section 3.1, we propose the multistep test statistic:

$$T_n \equiv \min_{c \in \mathscr{C}_n, R\pi = r} \left\| \sqrt{n} G_{n,c}\left(\pi\right) \right\|^2_{W_n(c)}, \tag{4.2}$$

where $\mathscr{C}_n$ is the effective parameter space for c_0 in the last step of the MMS.[18] See Section 3.1.

4.1 Asymptotic Validity and Consistency

Note that we are dealing with a post-selection inference problem because of the built-in moment selection in our test statistic. Albeit the many challenges faced with general post-selection inference documented in Leeb and Pötscher (2005) and Leeb

[18] A test analogous to the J-test can be applied to testing the joint validity of the model assumptions including the parametric form of the payoff function such as that in Game 1. Details are omitted due to space considerations.

and Pötscher (2008), we are able to construct an asymptotically uniformly valid and computationally simple test based on T_n.

Assume that the model is fully characterized by $\xi \in \Xi$, where Ξ is some compact parameter space and is possibly infinite dimensional. For the Simple Game, ξ includes the distribution of private information, conditional probability of latent state on observed state, and payoff vectors for each individual. Denote Ξ_R as the parameter space consistent with the null hypothesis and $\Pr_\xi(\cdot)$ as the probability calculated under ξ. The objective is to find a critical value CV that controls the asymptotic size defined as:

$$AsySize \equiv \limsup_{n\to\infty} \sup_{\xi\in\Xi_R} \Pr_\xi(T_n > CV). \tag{4.3}$$

We consider the drifting model parameter sequence ξ_n and the set of drifting model parameter sequences under H_0 with limit ξ as

$$\Xi_R(\xi) = \{\{\xi_n \in \Xi_R : n \geq 1\} : \xi_n \to \xi \in \Xi_R\}. \tag{4.4}$$

The important role of the analysis under drifting (sub)sequences has been emphasized in Andrews and Cheng (2012), Cheng (2015), and Andrews et al. (2020). Its introduction is not intended as a literal description of real-world data, but merely a device that helps us study the asymptotic property of the test statistic that mimics its finite-sample behavior. For the system of moment functions $G(\pi)$, a selection is correct if $G_c(\pi) = \mathbf{0}$ for some π, and is incorrect if $G_c(\pi) \neq \mathbf{0}$ for any π. By Lemma 2.1, the only true selection is c_0. However, under drifting sequence of model parameters, it is possible that $G_c^{\xi_n}(\pi) \neq \mathbf{0}$ but $G_c^{\xi_n}(\pi) \to \mathbf{0}$ for some $c \neq c_0$ and $\pi \in \Pi$ as $n \to \infty$, where $G_c^{\xi_n}(\pi)$ denotes the moment functions selected by c under the drifting model parameter sequence ξ_n. We call such selection a *nearly true selection*. When nearly true selections exist, the probability that c_0 is the solution to the minimization problem (4.2) does not approach one, so that incorrect moment functions may be selected even when $n \to \infty$. Such phenomenon occurs in post-selection inference problem and often complicates the inference procedure. However, since the number of correct moments in the Simple Game is known to be l, the null asymptotic distribution of T_n under drifting sequence is stochastically dominated by the chi-squared distribution with $(l - l_\pi + l_R)$-degrees of freedom. This allows us to construct an asymptotically uniformly valid test using critical value from the chi-squared distribution with $(l - l_\pi + l_R)$-degrees of freedom even in the presence of nearly true selections. Moreover, by Assumption 2.6, for any $c \neq c_0$, the system $G_c(\pi) = \mathbf{0}$

does not have a solution. The test statistic diverges to infinity under the alternative hypothesis because $\lim_{n\to\infty} \min_{R\pi=r} \|G_{n,c}(\pi)\|^2_{W_n(c)} > 0$ for all $c \in \mathscr{C}$ if $R\pi_0 \neq r$.

Assumption 4.1. *(i) The derivative of $f(\cdot)$ is bounded. (ii) For any $\xi \in \Xi_R$, $\mathbf{z}$ takes each value in $\mathscr{Z}$ with probability bounded below by $\varepsilon > 0$. (iii) For any $\xi \in \Xi_R$ and the parameter sequence $\{\xi_n\} \in \Xi_R(\xi)$, given each $c \in \mathscr{C}_n$, $W_n(c) = W(c) + o_p(1)$ with $W(c)$ being positive definite. (iv) $W(c_0) = \Omega_0^{-1}$ for Ω_0 being the asymptotic variance of $\sqrt{n}(G_{n,c_0}(\pi_0) - G_{c_0}(\pi_0))$.*

Assumption 4.1 (i) is satisfied by commonly used distributions and is needed for the uniform linear representation of the moment function given a uniform linear representation of the CCP estimator. Assumption 4.1 (ii) assumes that the support of $\mathbf{z}$ is the same for all $\xi \in \Xi_R$ and is one of the sufficient conditions needed for the uniform linear representation of Xiao (2018)'s CCP estimator. Assumption 4.1 (iii) and (iv) require that the probability limit of $W_n(c)$ be positive definite and that the optimal weighting matrix be used for c_0. The asymptotic variance matrix Ω_0 being nonsingular is satisfied automatically by the setting of the Simple Game and is verified in the proof Theorem 4.1.

Denote $\chi^2_{[df],1-\alpha}$ as the $(1-\alpha)$-th quantile of the chi-squared distribution with df-degrees of freedom. The following theorem states the asymptotic validity and consistency of the test based upon the test statistic T_n and critical value $\chi^2_{[l-l_\pi+l_R],1-\alpha}$. The proof is provided in the online appendix.

Theorem 4.1. *Let Assumptions 2.1-2.6 hold. For any $l_1 \in \{l_\pi, l_\pi+1, .., l\}$, $\alpha_1 \in (0,1]$, $\lambda \in (-1,0)$, and $\Delta \in \{1, .., l-l_1\}$, (i) if in addition Assumptions 3.1 and 4.1 hold, then*

$$\limsup_{n\to\infty} \sup_{\xi\in\Xi_R} \Pr_\xi\left(T_n > \chi^2_{[l-l_\pi+l_R],1-\alpha}\right) = \alpha;$$

(ii) if in addition Assumptions 3.1-3.2 hold, then for any $\xi \notin \Xi_R$,

$$\lim_{n\to\infty} \Pr_\xi\left(T_n > \chi^2_{[l-l_\pi+l_R],1-\alpha}\right) = 1.$$

Remark 4.1. We can also test hypotheses involving cross-player restrictions on some latent state or cross-latent state linear restrictions for some player. Both can be achieved by stacking the moment functions and adjusting the parameter space of the true selection vector. For example, let the null hypothesis be that the payoffs for players 1 and 2 are equal on some latent state. For $i = 1, 2$, we can construct the

sample moment functions $G_{ni}(\pi_i) = \overline{\pi}_{ni} - \Gamma_{ni}\pi_i$ for player i. Define

$$G_n(\pi_1, \pi_2) \equiv \begin{bmatrix} G_{n1}(\pi_1) \\ G_{n2}(\pi_2) \end{bmatrix} = \begin{bmatrix} \overline{\pi}_{n1} \\ \overline{\pi}_{n2} \end{bmatrix} - \begin{bmatrix} \Gamma_{n1} & \mathbf{0} \\ \mathbf{0} & \Gamma_{n2} \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} \text{ and}$$
$$\mathscr{C} \equiv \left\{ [c_1, \ldots, c_{2l}]^\top \in \mathbb{R}^{4l} : c_t = c_{t+l} \in \{[1,0], [0,1]\} \text{ for } t \in \{1, ..., l\} \right\}.$$

In the construction of $\mathscr{C}$, we keep the ordering of mixing components the same across players because the CCPs for different players are identifiable up to the same label swapping. The test can be carried out by the test statistic

$$\min_{c \in \mathscr{C}_n, \pi_1 = \pi_2} \left\| \sqrt{n} G_{n,c}(\pi_1, \pi_2) \right\|^2_{W_n(c)}$$

and the critical value $\chi^2_{[2l-l_\pi],1-\alpha}$, where $\mathscr{C}_n$ is the effective parameter space constructed from a similar MMS procedure.

4.2 Bootstrap Estimation of the Weighting Matrix

To implement the proposed test for the Simple Game, the weighing matrix needs to satisfy Assumption 4.1 (iii) and (iv), which involves estimating the asymptotic covariance matrix of $\sqrt{n}(G_{nc_0}(\pi_0) - G_{c_0}(\pi_0))$. Since $\overline{\pi}_n$ and Γ_n are obtained from plugging in estimators of the equilibrium CCPs via the eigendecomposition procedure, estimating Ω_0 from its analytical expression can be difficult. We propose a nonparametric bootstrap estimator of Ω_0 in this section.

Any $\pi \in \Pi$ satisfying the system of linear equations $R\pi = r$ can be expressed as $\Psi\pi_f + \mu$, where Ψ is a known $l_\pi \times (l_\pi - l_R)$ matrix, π_f is the free parameter vector of dimension $l_\pi - l_R$, and μ is a known $l_\pi \times 1$ vector. Computation of bootstrap weighting matrix $W_n^b(c)$ includes the following steps.

Step 1: For any given $c \in \mathscr{C}_n$, if $\arg\min_{\pi_f} \|G_{n,c}(\Psi\pi_f + \mu)\|^2$ is not unique, then set $W_n^b(c)$ as some known positive definite matrix W_P such as the identify matrix. Otherwise, let $\widehat{\pi}_f(c) = \arg\min_{\pi_f} \|G_{n,c}(\Psi\pi_f + \mu)\|^2$ and continue to Step 2.

Step 2: Compute the bootstrap variance

$$\Sigma_n^b(c, \widehat{\pi}_f(c)) = \frac{n}{B} \sum_{b=1}^{B} \left(G_{n,c}^{(b)}(\Psi\widehat{\pi}_f(c) + \mu) - \overline{G}_{n,c} \right) \left(G_{n,c}^{(b)}(\Psi\widehat{\pi}_f(c) + \mu) - \overline{G}_{n,c} \right)^\top,$$

where $G_{n,c}^{(b)}(\cdot)$ is calculated using the b-th nonparametric bootstrap sample and $\overline{G}_{n,c} = \frac{1}{B}\sum_{b=1}^{B} G_{n,c}^{(b)}(\Psi\widehat{\pi}_f(c) + \mu)$. Set $W_n^b(c)$ as $\left(\Sigma_n^b(c, \widehat{\pi}_f(c))\right)^{-1}$.

Step 3: Repeat Step 1 and Step 2 for every $c \in \mathscr{C}_n$.

By Assumption 2.6, $\text{rank}\left(\Gamma_{n,c_0}\Psi\right) = l_\pi - l_R$ for n sufficiently large. However, for $c \neq c_0$, it is possible that $\text{rank}\left(\Gamma_c\Psi\right) < l_\pi - l_R$ and $\arg\min_{\pi_f} \left\|G_{n,c}\left(\Psi\pi_f + \mu\right)\right\|^2$ is not unique. Fortunately, for such c, Assumption 4.1 (iii) only requires the weighing matrix to be positive definite in the limit, which is satisfied by matrix W_P.

For $\Sigma_n^b\left(c_0, \widehat{\pi}_f\left(c_0\right)\right)$ to be a consistent estimator of Ω_0, we need to match the labels of latent states across different bootstrap draws. By Assumption 2.5, $p_i\left(\mathbf{z}, k\right) \neq p_i\left(\mathbf{z}, k'\right)$ for any $\mathbf{z} \in \mathscr{Z}$ and $i = 1, 2, 3$. In consequence, the labels of latent states can be matched through $p_i\left(\mathbf{z}, k\right)$ and $p_i\left(\mathbf{z}, k'\right)$ across bootstrap draws with probability approaching 1 as n goes to infinity. The property of $W_n^b\left(c\right)$ is summarized in the following proposition. The online appendix contains the proof of the proposition.

Proposition 4.1. *Suppose Assumptions 2.1-2.6, 3.1, and 4.1 (i) and (ii) hold. Then $W_n^b\left(c\right)$ satisfies Assumption 4.1 (iii) and (iv) when $B \to \infty$.*

5 The General Game

In this section, we extend the methods of estimation and inference for the Simple Game developed in Sections 3 and 4 to the *General Game* studied in Aguirregabiria and Mira (2019). The General Game allows for a general number of players, a general number of actions for each player, and more importantly both unobserved heterogeneity and multiple equilibria. In Section 5.1, we review the sequential identification of the General Game. Sections 5.2 and 5.3 introduce the estimation and inference procedures. Asymptotic results including assumptions and theorems are relegated to Appendix A.

5.1 The General Game and Sequential Identification

In the General Game, there are N players; player i has $J+1$ available actions denoted as $d_i \in \{0, ..., J\}$; there is an exclusive observed state variable $z_i \in \mathscr{Z}_i$ and one payoff-relevant state variable denoted as $k \in \mathscr{K}$ that is common knowledge for players but unobserved by the econometrician. The objective is to identify, estimate, and conduct inference on the payoff vector on each observed and latent state.

Let $\mathbf{z} \equiv \left(z_1, ..., z_N\right) \in \mathscr{Z} \equiv \prod_{i=1}^N \mathscr{Z}_i$ and $p^k\left(\mathbf{z}\right)$ be the conditional probability mass function of k given $\mathbf{z}$ with support $\mathscr{K}$. To account for potential multiple equilibria in the data, we introduce a bivariate injective function denoted as $b\left(\cdot, \cdot\right)$ and define

$\omega \equiv b(k,\tau)$ as the scalar *composite latent variable* for the combination of latent state k and equilibrium indicator τ. Let $p(\omega \mid \mathbf{z})$ be the conditional probability mass function of ω given $\mathbf{z}$ with support $\Omega_{\mathbf{z}}$. For any given $\mathbf{z}$, $|\Omega_{\mathbf{z}}|$ is unknown but can be identified using existing methods such as Lemma 1 in Xiao (2018) under Assumption C.1 provided in the online appendix of this paper. Without loss of generality, we assume that $\omega \in \{1, ..., |\Omega_{\mathbf{z}}|\}$ for any $\mathbf{z} \in \mathscr{Z}$. In the presence of multiple equilibria, the support of the composite latent variable ω depends on $\mathbf{z}$, because the number of active equilibria might change with $\mathbf{z}$.

Let $\boldsymbol{\epsilon}_i \equiv (\epsilon_i(1), \ldots, \epsilon_i(J))^\top$, where $\epsilon_i(j)$ is the normalized (with respect to action 0) private information for player i taking action j. We impose the following assumptions on the private information, data, and latent states that are analogous to Assumptions 2.1-2.4 (i).

Assumption 5.1. *(i)* $\{\boldsymbol{\epsilon}_i\}_{i=1}^N \overset{i.i.d}{\sim} F(\cdot)$*, where* $F(\cdot)$ *is known to the econometrician and is absolutely continuous with a probability density function denoted as* $f(\cdot)$*. (ii) The support of* $f(\cdot)$ *is* $\mathbb{R}^J$*. (iii)* $\{\boldsymbol{\epsilon}_i\}_{i=1}^N$ *are independent of the state variables* $(\mathbf{z}, k)$*. (iv) The econometrician observes a random sample* $\{(d_{1m}, \ldots, d_{Nm}, z_{1m}, \ldots, z_{Nm})\}_{m=1}^n$*. (v)* $p^k(\mathbf{z}) > 0$ *for any* $k \in \mathscr{K}$ *and* $\mathbf{z} \in \mathscr{Z}$*.*

Different from the Simple Game, neither the number of multiple equilibria nor the number of latent states $|\mathscr{K}|$ is known in the General Game. Like for the Simple Game, we focus our discussion on sequential identification and estimation of the payoff vectors for player 1 with $z_1 = z_1^1$ fixed, where z_1^1 is the first element in $\mathscr{Z}_1$. Holding $z_1 = z_1^1$, define $l \equiv \prod_{i=2}^N |\mathscr{Z}_i|$ as the total number of different observed states. When $|\mathscr{Z}_i|$ is the same for all $i = 2, \ldots, N$, l is an exponential function of the number of players. Let $\{\mathbf{z}^1, \ldots, \mathbf{z}^l\}$ be the collection of all the observed states.

Sequential identification of the payoff vectors consists of two steps. Because of multiple equilibria, the composite latent variable ω plays the same role as the unobserved heterogeneity k in the Simple Game. So in the first step, we identify $|\Omega_{\mathbf{z}}|$ and the CCPs up to a label swapping in ω. In the second step, we identify $|\mathscr{K}|$ and the payoff vector on the latent variable k by first extending Step-2 for the Simple Game to identify the payoff vector on ω and then identifying the payoff vectors on the latent variable k from those on ω.

5.1.1 Step-1: Identification of the Equilibrium CCP Vector (up to a Label Swapping in the Composite Latent Variable)

For a given $\mathbf{z}$, identification of equilibrium CCPs for the General Game is equivalent to the identification of the following nonparametric mixture model with $|\Omega_{\mathbf{z}}|$ number of mixing components: for $(d_1, ..., d_N) \in \{0, ..., J\}^N$,

$$p(d_1, ..., d_N \mid \mathbf{z}) = \sum_{\omega \in \Omega_{\mathbf{z}}} \left[p(\omega \mid \mathbf{z}) \prod_{i=1}^{N} p(d_i \mid \mathbf{z}, \omega) \right]. \tag{5.1}$$

One can apply the approach in either Bonhomme et al. (2016) or Xiao (2018). Instead of restating these conditions here, we simply assume Step-1 identification. In Appendix C.2, we provide a detailed analysis of the approach in Xiao (2018).

Assumption 5.2. *(i) For any* $\mathbf{z} \in \{\mathbf{z}^1, \ldots, \mathbf{z}^l\}$, $|\Omega_{\mathbf{z}}|$ *is identified. (ii) Equilibrium CCPs are identified up to a label swapping in* ω.

Since $|\Omega_{\mathbf{z}}| \geq |\mathscr{K}|$ and the inequality is strict when there are multiple equilibria, Step-1 does not identify $|\mathscr{K}|$ which will be done in Step-2.

5.1.2 Step-2: Identification of the Payoff Vector

Denote the expected payoffs and the identified CCPs for all actions on some $(\mathbf{z}, \omega)$ as

$$\begin{aligned} \overline{\pi}_1(\mathbf{z}, \omega) &\equiv [\overline{\pi}_1(1, \mathbf{z}, \omega), \ldots, \overline{\pi}_1(J, \mathbf{z}, \omega)]^\top \text{ and} \\ \mathbf{p}_1(\mathbf{z}, \omega) &\equiv [\Pr(d_1 = 1 \mid \mathbf{z}, \omega), \ldots, \Pr(d_1 = J \mid \mathbf{z}, \omega)]^\top. \end{aligned}$$

Further let the mapping from equilibrium expected payoffs to CCPs be:

$$\Lambda_{1\mathbf{z}\omega}(\overline{\pi}_1(\mathbf{z}, \omega)) = \begin{bmatrix} \Pr(\overline{\pi}_1(1, \mathbf{z}, \omega) + \epsilon_1(1) \geq \overline{\pi}_1(j, \mathbf{z}, \omega) + \epsilon_1(j) \text{ for every } j) \\ \vdots \\ \Pr(\overline{\pi}_1(J, \mathbf{z}, \omega) + \epsilon_1(J) \geq \overline{\pi}_1(j, \mathbf{z}, \omega) + \epsilon_1(j) \text{ for every } j) \end{bmatrix} = \mathbf{p}_1(\mathbf{z}, \omega).$$

The inverse of this mapping $\Lambda^{-1}_{1\mathbf{z}\omega}(\mathbf{p}_1(\mathbf{z}, \omega))$ identifies $\overline{\pi}_1(\mathbf{z}, \omega)$ from CCPs (see Hotz and Miller (1993) and Aguirregabiria and Mira (2019)). Given the identification of expected payoffs for all $(\mathbf{z}, \omega)$, we construct player 1's expected payoff vector $\overline{\pi}$ and coefficient matrix Γ for $z_1 = z_1^1$ by stacking the corresponding expected payoffs and CCPs for all actions, $\mathbf{z}$ and ω. See Appendix A.1 for their expressions.

Given $\overline{\pi}$ and Γ, we construct the system of moment functions: $G(\pi) = \overline{\pi} - \Gamma\pi$, which will be used to identify the true payoff vectors denoted as $t\pi_0^k$ for $k = 1, \ldots, |\mathscr{K}|$,

where π has dimension $l_\pi \equiv J\left(J+1\right)^{N-1}$. Identification of $t\pi_0^k$ for $k=1,\ldots,|\mathscr{K}|$ is complicated by the presence of multiple equilibria. In contrast to the Simple Game, Step-1 in the General Game only identifies equilibrium CCPs on the composite latent variable ω up to a label swapping instead of equilibrium CCPs on the latent variable k in the Simple Game. Step-2 identification for the Simple Game is therefore adapted to first identify the payoff vector π_0^h on the composite latent variable for $h=1,...,|\Omega_{\mathbf{z}^1}|$ in Step-2(i) below and then to include an additional step to identify $t\pi_0^k$ for $k=1,\ldots,|\mathscr{K}|$ based on the identified π_0^h for $h=1,...,|\Omega_{\mathbf{z}^1}|$ in Step-2(ii).

Step-2(i). Similar to Step-2 for the Simple Game, we first select all the equilibrium CCPs corresponding to the same latent state across observed states. Extending Definition 2.3 for the Simple Game, a correct matching should select all components of $G\left(\cdot\right)$ that correspond to the same latent state. Each matching can be represented by a selection vector of zeros and ones, where the true selection vector corresponds to the correct matching. We construct $|\Omega_{\mathbf{z}^1}|$ number of parameter spaces which is the number of support points of ω conditioning on $\mathbf{z}^1$. Let $\mathbf{e}_1=[1,\ldots,1]^\top\in\mathbb{R}^J$ and $\mathbf{e}_0=[0,\ldots,0]^\top\in\mathbb{R}^J$. The parameter spaces of true selection vectors are defined in the following.

Definition 5.1. *For $h=1,...,|\Omega_{\mathbf{z}^1}|$, the parameter space of true selection vectors is defined as*

$$\mathscr{C}^h\equiv\left\{\begin{array}{c}[c_1,...,c_l]^\top : c_1=\left[c_{1,1},\ldots,c_{1,|\Omega_{\mathbf{z}^1}|}\right] \text{ with } c_{1,h}=\mathbf{e}_1 \text{ and } c_{1,j}\in\{\mathbf{e}_1,\mathbf{e}_0\}\\ \text{for } j\neq h;\ \text{for } t=2,\ldots,l,\ c_t=\left[c_{t,1},\ldots,c_{t,|\Omega_{\mathbf{z}^t}|}\right] \text{ with } c_{t,\mathrm{w}}\in\{\mathbf{e}_1,\mathbf{e}_0\},\\ \text{where } \mathrm{w}\in\{1,\ldots,|\Omega_{\mathbf{z}^t}|\},c_t\neq[\mathbf{e}_0,\ldots,\mathbf{e}_0]\end{array}\right\}.$$

For each $h=1,...,|\Omega_{\mathbf{z}^1}|$, $\mathscr{C}^h$ contains $2^{|\Omega_{\mathbf{z}^1}|-1}\prod_{t=2}^{l}\left(2^{|\Omega_{\mathbf{z}^t}|}-1\right)$ number of elements, where each element has dimension $J\sum_{t=1}^{l}|\Omega_{\mathbf{z}^t}|$. Every selection vector in $\mathscr{C}^h$ selects the h-th mixing component on $\mathbf{z}^1$. Because the h-th mixing component corresponds to just one latent state, $\mathscr{C}^h$ contains only one true selection vector denoted as c_0^h. The parameter spaces of true selection vectors in the General Game differ from those in the Simple Game in several ways. First, a selection vector in $\mathscr{C}^h$ is able to select more than one mixing component per observed state to accommodate the possible presence of multiple equilibria. Second, there might be more than one selection vector in $\mathscr{C}^h$ for any $h=1,...,|\Omega_{\mathbf{z}^1}|$ that generates a system with a solution (not necessarily unique). The reason is that as long as the components of $G\left(\cdot\right)$ selected by c correspond to the same latent state, $G_c\left(\pi\right)=\mathbf{0}$ has at least the true payoff vector

as its solution. Third, because the h_1-th and h_2-th mixing components on $\mathbf{z}^1$ can correspond to the same latent state, we can have $c_0^{h_1} = c_0^{h_2}$ even if $h_1 \neq h_2$.

Define a set

$$\mathscr{C}\mathscr{I}^h \equiv \left\{ c \in \mathscr{C}^h : G_c(\pi) = \mathbf{0} \text{ has a solution} \right\}.$$

The true selection vector c_0^h differs from other elements in $\mathscr{C}\mathscr{I}^h$ by selecting *all* the equilibria per observed state that correspond to the same latent state. As a result, c_0^h is the unique element in $\mathscr{C}\mathscr{I}^h$ that selects the maximum number of components of $G(\cdot)$ and π_0^h is the unique solution to $G_{c_0^h}(\pi) = \mathbf{0}$.

Step-2(ii). Based on $\pi_0^1, \ldots, \pi_0^{|\Omega_{\mathbf{z}^1}|}$, we can identify $|\mathscr{K}|$ and $t\pi_0^k$ for $k = 1, \ldots, |\mathscr{K}|$ as the distinct elements among $\pi_0^1, \ldots, \pi_0^{|\Omega_{\mathbf{z}^1}|}$.

Similar to Assumption 2.6 for the Simple Game, we adopt the following assumption for Step-2 identification.

Assumption 5.3. *For each $h = 1, ..., |\Omega_{\mathbf{z}^1}|$, (i) $\Gamma_{c_0^h}$ has full column rank and (ii)* $\text{rank}([\overline{\pi}_c, \Gamma_c]) > \text{rank}(\Gamma_c)$ *for any $c \in \mathscr{C}^h$ that selects different latent states.*

Assumption 5.3 is equivalent to the necessary and sufficient condition proposed in Proposition 3 of Aguirregabiria and Mira (2019) except that we consider identification for each pair of player and exclusive state separately. See Appendix C.4 for more discussion on the equivalence result.

We summarize the identification results in Step-2 in the following lemma.

Lemma 5.1. *Let Assumptions 5.1-5.3 holds. (i) For $h = 1, ..., |\Omega_{\mathbf{z}^1}|$, $\left(c_0^h, \pi_0^h\right)$ is identified: $\left\|c_0^h\right\|_0 > \|c\|_0$ for any $c \in \mathscr{C}\mathscr{I}^h$ and $c \neq c_0^h$; and $G_{c_0^h}(\pi) = \mathbf{0}$ has a unique solution π_0^h. (ii) $|\mathscr{K}|$ is identified as the number of distinct payoff vectors among $\pi_0^1, \ldots, \pi_0^{|\Omega_{\mathbf{z}^1}|}$ and $t\pi_0^k$ for $k = 1, \ldots, |\mathscr{K}|$ are identified as the $|\mathscr{K}|$ distinct payoff vectors.*

5.2 Estimation

It follows from Lemma 5.1 (ii) that estimation of $|\mathscr{K}|$ and $t\pi_0^k$ for $k = 1, \ldots, |\mathscr{K}|$ requires an initial estimator of π_0^h for $h = 1, ..., |\Omega_{\mathbf{z}^1}|$. In the following, we first introduce an estimation procedure for π_0^h that extends the MMS procedure for the Simple Game. Then we propose an estimator for $|\mathscr{K}|$ and $t\pi_0^k$. In Appendix A.3, we show that our estimators are consistent. Additionally, we prove that when the number of the observed states where multiple equilibria exist is not a function of l, the time and space complexities of our estimation procedure achieve linearity in l.

5.2.1 Multistep Moment Selection Estimation of π_0^h

To construct an estimator of $\left(c_0^h, \pi_0^h\right)$ based on the identification strategy in Lemma 5.1 (i), we can follow the idea in Andrews (1999) and define the estimator as

$$\left(\widetilde{c}^h, \widetilde{\pi}^h\right) = \arg \min_{c \in \mathscr{C}^h, \pi \in \Pi} \left[\|G_{n,c}(\pi)\|^2_{W_n(c)} - \rho_1 \left(\|c\|_0\right) \kappa_{1,n}/n \right],$$

where $W_n(c)$ is the sample weighting matrix, $\rho_1(\cdot) > 0$ is a known strictly increasing function, and $\kappa_{1,n} \to \infty$ with $\kappa_{1,n} = o(n)$. The additional term, $-\rho_1 \left(\|c\|_0\right) \kappa_{1,n}/n$, rewards selecting more moments. By the same logic as in Andrews (1999), we expect $\widetilde{c}^h$ to be equal to c_0^h with probability approaching one and $\widetilde{\pi}^h$ to consistently estimate π_0^h for $h = 1, ..., |\Omega_{\mathbf{z}^1}|$.

Similar to the Simple Game, calculating $\left(\widetilde{c}^h, \widetilde{\pi}^h\right)$ is computationally challenging for large l. In Appendix A.2, we extend the MMS procedure developed in Section 3.1 to overcome such a challenge. Denote the resulting estimators as $\left(\widehat{c}^h, \widehat{\pi}^h\right)$ for $h = 1, 2, ..., |\Omega_{\mathbf{z}^1}|$.

5.2.2 Estimation of $|\mathscr{K}|$ and $t\pi_0^k$

By Lemma 5.1 (ii), we identify $|\mathscr{K}|$ as the number of groups of π_0^h for $h = 1, ..., |\Omega_{\mathbf{z}^1}|$, such that π_0^h's are the same within a group but different across groups. Although $\widehat{\pi}^h$ obtained from the previous section is a consistent estimator, estimation error in finite sample will cause $\widehat{\pi}^{h_1}$ to differ from $\widehat{\pi}^{h_2}$ even if $\pi_0^{h_1} = \pi_0^{h_2}$. To obtain a grouping of the estimated payoffs by their underlying latent states in finite sample, we extend the idea in k-means clustering allowing for unknown number of clusters.

Given some integer $K \leq |\Omega_{\mathbf{z}^1}|$, denote $\mathbf{S}_K \equiv \{S_{K,1}, \ldots, S_{K,K}\} \in \mathfrak{S}_K$ as a partition of $\left\{\widehat{\pi}^1, \ldots, \widehat{\pi}^{|\Omega_{\mathbf{z}^1}|}\right\}$ with cardinality K, where $\mathfrak{S}_K$ denotes the collection of all partitions with cardinality K. Let $\mu_K \equiv \{\mu_{K,1}, \ldots, \mu_{K,K}\}$, where for $j = 1, \ldots, K$, $\mu_{K,j} \equiv \frac{1}{|S_{K,j}|} \sum_{\widehat{\pi}^h \in S_{K,j}} \widehat{\pi}^h$ is the average of $\widehat{\pi}^h$'s within the j-th group in the partition. Let $\rho_2(\cdot) > 0$ be an increasing function and $\kappa_{2,n} \to \infty$ with $\kappa_{2,n} = o(n)$. We estimate $|\mathscr{K}|$ and $t\pi_0^k$ by solving the following minimization problem:

$$\left(\widehat{K}, \widehat{\mathbf{S}}_{\widehat{K}}\right) = \arg \min_{K \leq |\Omega_{\mathbf{z}^1}|, \mathbf{S}_K \in \mathfrak{S}_K} \left[\sum_{j=1}^{K} \sum_{\widehat{\pi}^h \in S_{K,j}} \left\|\widehat{\pi}^h - \mu_{K,j}\right\|^2 + \rho_2(K) \kappa_{2,n}/n \right]. \quad (5.2)$$

The solution $\widehat{K}$ is the estimator for $|\mathscr{K}|$. Let $\widehat{t\pi}^1, \ldots, \widehat{t\pi}^{\widehat{K}}$ be the within-group averages of $\widehat{\pi}^h$ that correspond to partition $\widehat{\mathbf{S}}_{\widehat{K}}$. Then $\widehat{t\pi}^1, \ldots, \widehat{t\pi}^{\widehat{K}}$ are the estimators for

the true payoff vectors.

Comparing (5.2) with the objective function in the classical k-means clustering, (5.2) includes an additional term $\rho_2(K)\kappa_{2,n}/n$ that penalizes large K. To see the intuition behind this objective function, first note that if $K < |\mathscr{K}|$, the correct partition is not feasible; and the first term $\sum_{j=1}^{K}\sum_{\widehat{\pi}^h\in S_{K,j}}\left\|\widehat{\pi}^h-\mu_{K,j}\right\|^2$ does not converge in probability to zero under any partition. When $K \geq |\mathscr{K}|$, the correct partition is feasible. The first term converges in probability to zero under the correct partition or its refinements.[19] The penalty term $\rho_2(K)\kappa_{2,n}/n$ makes the objective function to favor partitions that use smaller K. As a result, both events $\widehat{K} = |\mathscr{K}|$ and $\widehat{\mathbf{S}}_{\widehat{K}}$ equals to the correct partition occur with probability approaching one as n goes to infinity. Given the correct partition, the group centers $\widehat{t\pi}^1,\ldots,\widehat{t\pi}^{\widehat{K}}$ deliver consistent estimate for the true payoff vectors on each latent state.

The value of $|\Omega_{\mathbf{z}^1}|$ is usually small (no more than 10, see Igami and Yang (2016) and Bonhomme et al. (2022)). As a result, applying the k-means to classify $\left\{\widehat{\pi}^1,\ldots,\widehat{\pi}^{|\Omega_{\mathbf{z}^1}|}\right\}$ is fast; and the estimators $\left(\widehat{K},\widehat{\mathbf{S}}_{\widehat{K}}\right)$ can be solved within a millisecond.[20] In addition, the running time does not depend on the sample size or the number of moments l.

5.3 Inference

Researchers may want to test linear restrictions on the payoff vector that associates with the h-th mixing component on $\mathbf{z}^1$. Such hypotheses can be expresses as

$$H_0 : R\pi_0^h = r \text{ against } H_1 : R\pi_0^h \neq r,$$

where R is a known matrix of dimension $l_R \times l_\pi$ with $\text{rank}(R) = l_R$ and r is a known vector of dimension $l_R \times 1$. Similar to the Simple Game, the test statistic

$$\min_{c\in\mathscr{C}^h, R\pi=r}\left\|\sqrt{n}G_{n,c}(\pi)\right\|^2_{W_n(c)} \tag{5.3}$$

can be used. However, because (5.3) is difficult to compute for large l, we exploit the MMS procedure developed in Appendix A.2. By replacing $\mathscr{C}^h$ with the effective

[19] A partition S' is a refinement of a partition S if every element of S' is a subset of some element of S.

[20] In rare cases where numbers of mixing components are large among all the observed states, existing algorithms such as Lloyd's algorithm (Lloyd (1982)) can be applied to improve the computational time. Unfortunately, these algorithms are not exact and are subject to multiple optima issues.

parameter space $\mathscr{C}_n^h$, we propose the following test statistic:

$$T_n \equiv \min_{c \in \mathscr{C}_n^h, R\pi = r} \left\| \sqrt{n} G_{n,c}(\pi) \right\|_{W_n(c)}^2 . \tag{5.4}$$

The size of $\mathscr{C}_n^h$ is much smaller than that of $\mathscr{C}^h$, making T_n fast to compute. In Appendix A.4, we prove that the proposed test has the correct asymptotically size and is consistent.

In some cases, we may wish to test if any of the true payoff vectors satisfies a linear restriction:

$$\begin{aligned} &H_0 : R(t\pi_0^k) = r \text{ for some } k = 1, \ldots, |\mathscr{K}| \text{ against} \\ &H_1 : R(t\pi_0^k) \neq r \text{ for all } k = 1, \ldots, |\mathscr{K}| . \end{aligned}$$

A computationally feasible test statistic defined as

$$\min_{c \in \mathscr{C}_n^1 \cup \cdots \cup \mathscr{C}_n^{\left|\Omega_{\mathbf{z}^1}\right|}, R\pi = r} \left\| \sqrt{n} G_{n,c}(\pi) \right\|_{W_n(c)}^2$$

can be applied and it behaves similarly to T_n defined in (5.4).

6 Monte Carlo Simulation

In this section, we conduct a Monte Carlo simulation study to examine the performance of our estimation and inference methods in finite samples. We use two variants of the Simple Game and two variants of the General Game in this study to illustrate the applicability of the proposed methods beyond the Simple Game and the General Game. The two variants of the Simple Game are called Game 1 and Game 2; and the two variants of the General Game are called Game 3 and Game 4, respectively. In all four games, we fully parametrize the payoff functions. Game 1 and Game 2 are games with only unobserved heterogeneity but no multiple equilibria. Game 3 and Game 4 are games with both unobserved heterogeneity and multiple equilibria. Game 1 has been introduced in Section 2.4. We use it to study the effect of the tuning parameters on running time and correct selection rate (CSR) of the MMS procedure, where the CSR is computed as the number of times that $\widehat{c} = c_0$ divided by the number of repetitions. Based on the simulation result, we recommend a rule-of-thumb for choosing the tuning parameters. Game 2 adds a common observed state to the Simple Game and a strategic effect that varies with state variables. It is used as a robustness check for

the effectiveness of the rule-of-thumb. We investigate the finite sample performance of the estimator and test. Game 3 and Game 4 are used to evaluate the performance of the MMS procedure together with the consistent grouping method that separate multiple equilibria from unobserved heterogeneity. All the results on the running time in this section are obtained from a computer of 2.4GHz CPU and 1TB RAM.

In Game 1, we let the normalized private information ϵ_i follow a logistic distribution. The structures of Games 2-4 are presented below.[21] The parameter values are provided in Appendix D.2.

Game 2: We consider a game with a common observed state variable $x \in \mathscr{X}$ that takes discrete values. Let the payoff function for player i choosing $d_i = 1$ be

$$\pi_i\left(1, \mathbf{d}_{-i}, z_i, x, k\right) = \beta_{ik} x + \delta_{ik} z_i \left(\sum_{j \neq i} d_j\right),$$

where $(\beta_{ik}, \delta_{ik})$ are the parameters of interest, β_{ik} characterizes the effect of x, and $\delta_{ik} z_i$ captures the strategic effect that varies with z_i. Both effects change with k. The normalized private information ϵ_i follows the standard normal distribution. The conditional distribution of the latent state is specified as:

$$\Pr\left(k = A \mid \mathbf{z}, x\right) = \begin{cases} \frac{3}{4} - \frac{1}{10\left(|x| + \sum_{i=1}^{3} |z_i|\right)} & \text{for } (\mathbf{z}, x) \neq \left(\mathbf{z}^\dagger, x^\dagger\right) \\ 0.45 & \text{for } (\mathbf{z}, x) = \left(\mathbf{z}^\dagger, x^\dagger\right) \end{cases},$$

for some $\mathbf{z}^\dagger \in \mathscr{Z}$ and $x^\dagger \in \mathscr{X}$. The values of $(\mathbf{z}, x)$ are chosen so that $\Pr\left(k = A \mid \mathbf{z}, x\right) > \Pr\left(k = B \mid \mathbf{z}, x\right)$ for all $(\mathbf{z}, x) \neq \left(\mathbf{z}^\dagger, x^\dagger\right)$. The ranking independence assumption does not hold, because the mixing weight for latent state A is not always strictly larger than that for latent state B.

Game 3: The game has 5 players, 18 observed states, and 2 latent states. The payoff function for player i when choosing $d_i = 1$ is given by

$$\pi_i\left(1, \mathbf{d}_{-i}, z_i, k\right) = x\theta_k + \delta_k \frac{1}{N-1} \left(\sum_{j \neq i} d_j\right),$$

where (θ_k, δ_k) are the parameters of interest. The normalized private information ϵ_i follows the standard normal distribution.

[21] Other models with similar parameters of interest include Example 1 in Kasahara and Shimotsu (2009), in which a single agent dynamic discrete choice model has two parameters that depend on a common latent states with finite support.

Game 4: The number of players, observed states, and latent states are the same as Game 3. We let the payoff function for player i choosing $d_i = 1$ be

$$\pi_i\left(1, \mathbf{d}_{-i}, z_i, k\right) = x\theta_k + \left(1 + x^2\right)\delta_k \frac{1}{N-1}\left(\sum_{j \neq i}\left(2d_j - 1\right)\right),$$

where (θ_k, δ_k) are the parameters of interest. This (normalized) payoff function is a result of the following two payoff functions for $d_i = 1$ and $d_i = 0$ respectively:

$$\widetilde{\pi}_i\left(1, \mathbf{d}_{-i}, z_i, k\right) = x\theta_k + \left(1 + x^2\right)\delta_k \frac{1}{N-1}\left(\sum_{j \neq i}\left(d_j = 1\right)\right) \text{ and}$$

$$\widetilde{\pi}_i\left(0, \mathbf{d}_{-i}, z_i, k\right) = \left(1 + x^2\right)\delta_k \frac{1}{N-1}\left(\sum_{j \neq i}\left(d_j = 0\right)\right).$$

The normalized private information ϵ_i follows the logistic distribution.

The identification strategies for Games 2-4 are similar to the one for Game 1 in Section 2.4. We discuss the identification and minimum-distance criterion of Games 2-4 in Appendix D.1.

6.1 Rule-of-Thumb Choice of Tuning Parameters in the MMS Procedure—Game 1

Before studying the effect of the tuning parameters in the MMS procedure, we demonstrate that Assumption 2.6 holds generically. We draw values for all primitive parameters independently from uniform distributions whose supports cover the parameter values used in the simulation. For all the drawn parameter values, Assumption 2.6 is satisfied. We interpret this as evidence that our identification assumption holds generically in the space of model primitives.

Table I: Effect of tuning parameters in the TMS: running time (in seconds) and correct selection rate (CSR)

		$l_1 = 8$		$l_1 = 9$		$l_1 = 10$		$l_1 = 11$		$l_1 = 12$		$l_1 = 13$	
α_1	λ	Time	CSR	Time	CSR	Time	CSR	Time	CSR	Time	CSR	Time	CSR
0.5%	−0.27	.0482	.93	.0417	.96	.0432	.99	.0598	.99	.0891	.99	.1529	.99
	−0.09	.0876	.93	.0397	.96	.0433	.99	.0596	.99	.0894	.99	.1524	.99
	−0.03	.1659	.94	.0762	.96	.0441	.99	.0594	.99	.0886	.99	.1524	.99
	−0.01	.2198	.96	.1082	.96	.0445	.99	.0607	.99	.0901	.99	.1537	.99
1%	−0.27	.0766	.95	.0666	.96	.0742	.99	.0820	.99	.1110	.99	.1764	.99
	−0.09	.0949	.95	.0689	.96	.0704	.99	.0816	.99	.1117	.99	.1770	.99
	−0.03	.1838	.95	.0961	.96	.0716	.99	.0817	.99	.1108	.99	.1760	.99
	−0.01	.2417	.97	.1170	.96	.0709	.99	.0823	.99	.1110	.99	.1742	.99
1.5%	−0.27	.0753	.95	.0948	.96	.0889	.99	.1046	.99	.1339	.99	.1993	.99
	−0.09	.0950	.95	.0900	.96	.0885	.99	.1057	.99	.1343	.99	.1992	.99
	−0.03	.1867	.95	.1135	.96	.0931	.99	.1047	.99	.1338	.99	.1989	.99
	−0.01	.2442	.97	.1529	.96	.0893	.99	.1043	.99	.1334	.99	.1985	.99
2.5%	−0.27	.1481	.95	.1369	.96	.1319	.99	.1509	.99	.1828	.99	.2514	.99
	−0.09	.1598	.95	.1329	.96	.1323	.99	.1507	.99	.1800	.99	.2511	.99
	−0.03	.2429	.95	.1505	.96	.1312	.99	.1504	.99	.1801	.99	.2509	.99
	−0.01	.2992	.97	.1679	.96	.1332	.99	.1496	.99	.1790	.99	.2505	.99

Table II: Effect of tuning parameters in the MMS: running time (in seconds) and correct selection rate (CSR)

			$\Delta = 1$		$\Delta = 2$		$\Delta = 3$		$\Delta = 4$		$\Delta = 5$		$\Delta = 6$	
l_1	α_1	λ	Time	CSR	Time	CSR	Time	CSR	Time	CSR	Time	CSR	Time	CSR
10	0.5%	−0.03	.0457	.99	.0423	.99	.0473	.99	.0529	.99	.0592	.99	.0729	.99
		−0.01	.0458	.99	.0428	.99	.0471	.99	.0514	.99	.0582	.99	.0727	.99
	1%	−0.03	.0531	.99	.0489	.99	.0546	.99	.0676	.99	.0821	.99	.1093	.99
		−0.01	.0527	.99	.0500	.99	.0574	.99	.0678	.99	.0840	.99	.1084	.99

			$\Delta = 1$		$\Delta = 2$		$\Delta = 3$		$\Delta = 4$		$\Delta = 5$		$\Delta = 6$	
l_1	α_1	λ	Time	CSR	Time	CSR	Time	CSR	Time	CSR	Time	CSR	Time	CSR
12	0.5%	−0.03	.1555	.99	.1564	.99	.1588	.99	.1953	.99	.2312	.99	.2639	.99
		−0.01	.1534	.99	.1585	.99	.1557	.99	.1945	.99	.2294	.99	.2618	.99
	1%	−0.03	.1771	.99	.1767	.99	.1922	.99	.2349	.99	.3078	.99	.3687	.99
		−0.01	.1748	.99	.1817	.99	.1875	.99	.2359	.99	.3092	.99	.3575	.99

Tables I and II study the effect of the tuning parameters in the TMS and MMS procedures on the running time and CSR. We use random samples with five hundred observations per observed state. The values in the table are based on one hundred

repetitions. For the design in Table I, $l = 18$; and for the design in Table II, $l = 27$. It can be seen from both tables that the tuning parameter l_1 plays an important role in reducing the running time and improving accuracy. As long as $l_1 \geq 10 = 5l_\pi$, the accuracy is higher or equal to 0.99. As we discussed in Section 3.3, increasing α_1 can increase the accuracy. But the effect is marginal. At the same time, the running time increases with α_1. When the accuracy is high, e.g. 99%, λ does not affect the computation time. Table II studies the role of Δ in the MMS on the running time and CSR. The results suggest that when the number of moments is large, a multistep procedure shall be applied. The estimator achieves a high CSR and short running time when $\Delta = 2$. When l increases from 18 to 27, the number of elements in the parameter space $\mathscr{C}$ increases more than five hundred times. However, because the novel MMS procedure is computationally very efficient, its running time only increases slightly.[22] As a comparison, $(\widetilde{c}, \widetilde{\pi})$ has an average running time around 3000 seconds when $l = 27$. Even with the least desirable choice of the tuning parameters, the MMS procedure is thousands of times faster than $(\widetilde{c}, \widetilde{\pi})$.

Based on Tables I and II, we recommend the following rule-of-thumb for setting the tuning parameters: $l_1 = 5l_\pi$, $\alpha_1 = 0.5\%$, $\lambda = -0.01$, and $\Delta = 2$. In the subsequent simulation study, we adopt this rule for all the games and designs.

To provide additional insight on the computational savings of the MMS procedure, we elaborate on its intermediate steps. Computing $(\widetilde{c}, \widetilde{\pi})$ requires solving a quadratic optimization problem $|\mathscr{C}|$ times; while calculating the MMS estimator $(\widehat{c}, \widehat{\pi})$ involves the same optimization problem $\sum_{s=1}^{S} |\mathscr{S}\mathscr{C}^s| + |\mathscr{C}_n|$ times. When $l = 18$, the cardinality of $\mathscr{C}$ is $2^{l-1} = 131072$. On the other hand, when implementing the MMS procedure with the rule-of-thumb proposed above, we have $S = 4$ with $\left|\mathscr{S}\mathscr{C}^1\right| = 512$ and $|\mathscr{S}\mathscr{C}^s| = |\mathscr{C}_n| = 12$ for $s = 2, 3, 4$ based on one simulation run. As a result, $(\widehat{c}, \widehat{\pi})$ requires only about 1/250 number of optimizations compared to $(\widetilde{c}, \widetilde{\pi})$.

Although $\mathscr{S}\mathscr{C}^1$ contains 512 sub-selection vectors, many are certainly incorrect. In consequence, the output of Step 1, $\mathscr{S}\mathscr{C}^1_n$, contains only 3 elements. More importantly, we do not just eliminate these $512 - 3 = 509$ number of sub-selection vectors, but all selection vectors that share the same first $2 \times l_1$ elements as the eliminated sub-selection vectors. In total, we eliminate $509 \times 2^{l-l_1} = 130304$ selection vectors after Step 1, leaving only a few selection vectors to be considered in the following steps.

[22] The running time for $l = 27$ can even be shorter than that for $l = 18$ because MMS rather than TMS is used when $l = 27$.

6.2 The MMS Procedure, Estimator, and Test Based on the Rule-of-Thumb

Applying the suggested rule-of-thumb for choosing the tuning parameters, we investigate the finite sample performance of the estimator and test in this section. For Game 1 and Game 2, we introduce five designs for each game corresponding to different numbers of observed states. The largest number of observed states considered in the simulation is set to be larger than the sizes of the observed states in studies such as Sweeting (2009), De Paula and Tang (2012), Grieco (2014), Igami and Yang (2016), and Xiao (2018).[23]

Table III: Finite sample performance of the MMS procedure in different designs: running time (in seconds), correct selection rate (CSR), and mean squared error (MSE)

		Design 1: $l = 18$			Design 2: $l = 27$			Design 3: $l = 64$			Design 4: $l = 100$			Design 5: $l = 288$		
	n_s	Time	CSR	MSE	Time	CSR	MSE	Time	CSR	MSE	Time	CSR	MSE	Time	CSR	MSE
Game 1	250	.0400	.944	.3824	.0448	.967	.0886	.1206	.912	.2142	.2338	.895	.3373	.4518	.795	.5273
	500	.0363	.996	.0749	.0433	.997	.0181	.0811	.994	.0355	.1980	.990	.0323	.4638	.982	.0750
	750	.0363	1	.0159	.0443	1	.0121	.0606	1	.0046	.1409	1	.0030	.3900	.998	.0162
	1000	.0325	1	.0121	.0436	1	.0086	.0591	1	.0033	.1198	1	.0020	.3737	1	.0012
		Design 1: $l = 24$			Design 2: $l = 36$			Design 3: $l = 54$			Design 4: $l = 81$			Design 5: $l = 162$		
	n_s	Time	CSR	MSE	Time	CSR	MSE	Time	CSR	MSE	Time	CSR	MSE	Time	CSR	MSE
Game 2	250	.0368	.933	.7600	.0307	.916	1.102	.0622	.845	2.366	.1992	.761	3.286	.4459	.740	3.596
	500	.0388	.996	.0725	.0313	.990	.1774	.0652	.980	.3602	.1658	.956	.7004	.3557	.956	.7468
	750	.0224	1	.0237	.0324	.998	.0623	.0615	.999	.0415	.1102	.991	.1656	.2742	.994	.1844
	1000	.0212	1	.0203	.0292	.999	.0205	.0531	1	.0144	.0839	.992	.1646	.2194	.996	.1422

Table III reports the running time, CSR, and mean squared error (MSE) of the MMS procedure for different sample sizes and different designs constructed from Games 1 and 2.[24] For each game, different designs correspond to different values

[23] In Sweeting (2009), De Paula and Tang (2012), and Xiao (2018), the major observed state is the rank of the market according to population, and there are 144 ranks in total; in Grieco (2014), the cardinality of observed states for the baseline model is 12 (3 status for discretized population, 2 status for whether it is active in 1998 and 2 status for whether there is a super-center within 20 miles); in the static game version of Igami and Yang (2016), the cardinality of observed states is 16 (4 categories for population and 4 categories for average income).

[24] We also run simulations with even smaller sample sizes. The estimated parameters become less accurate, whereas the running time is still short. As a sequential estimator, our MMS estimator is affected by the well-known finite sample bias documented in papers like Aguirregabiria and Mira (2007) and Aguirregabiria and Marcoux (2021). The bias largely results from poorly estimated CCPs at small sample sizes. Although developing CCP estimators with better finite sample performance

of l. n_s denotes the number of observations per observed state, and MSE reported in the table is calculated as the sum of MSEs of each parameter. The values in the table are computed from one thousand repetitions. Table III shows that the running time of the MMS increases very slowly with l. When the number of elements in the parameter space $\mathscr{C}$ increases millions of times, e.g., from $l = 27$ to $l = 64$ in Game 1, the running time of the MMS only increases about two times. Even when there are more than 10^{24} ($l = 81$) or 10^{29} ($l = 100$) matchings, the MMS takes less than one second to compute. This is consistent with Theorem 3.2, which shows that with probability approaching one, the time complexity of the MMS becomes linear in l when the sample size goes to infinity. The results in Table III also show that the running time either stays the same or decreases with the sample size. We suspect that the former case occurs because the running time is already close to being linear in l and has little room to improve with the sample size. When the sample size increases, the CSR increases and the MSE decreases. It would be ideal to compare the CSR and MSE of the MMS with $(\widetilde{c}, \widetilde{\pi})$. However, the extremely long running time of $(\widetilde{c}, \widetilde{\pi})$ makes the comparison impossible.

To study the finite sample performance of our inference procedure, we focus on Design 1 of Game 1 and consider two null hypotheses of the form $H_0 : R\pi_0 = r$ for $R = I_2$, $r = (\theta_{1A}, \delta_{1A})^\top$; and $R = (0, 1)$, $r = \delta_{1A}$. The first hypothesis is on the whole payoff parameter vector and the second one is on the parameter of strategic interaction, which is of great interest in empirical research. The number of observed states is $l = 18$. We set the nominal size as 5% and use the 95% quantile of χ^2_{18} and χ^2_{17} as the critical values for $R = I_2$ and $R = (0, 1)$. In both cases, the bootstrap weighting matrices are calculated with one thousand bootstrap samples. The results are based on five thousand Monte Carlo repetitions.

Table IV: Finite sample rejection probabilities under H_0 for different sample sizes

n_s	500	625	750	875	1000	1125	1250
$R = I_2$	0.0336	0.0334	0.0340	0.0394	0.0430	0.0406	0.0472
$R = (0, 1)$	0.0252	0.0262	0.0294	0.0350	0.0392	0.0386	0.0436

is beyond the scope of this paper, we see much value in future research in this direction.

Table V: Finite sample rejection probabilities under H_1 for different deviations and sample sizes

	Dev.	-0.15	-0.1	-0.05	-0.025	0.025	0.05	0.1	0.15
$R = I_2$	$n_s = 500$	0.9586	0.4704	0.0662	0.0338	0.0722	0.2004	0.7822	0.9950
	$n_s = 750$	0.9996	0.8298	0.1526	0.0422	0.0892	0.3012	0.9440	0.9998
	$n_s = 1000$	1	0.9608	0.2596	0.0672	0.1126	0.4228	0.9904	1
	Dev.	-0.2	-0.15	-0.1	-0.05	0.05	0.1	0.15	0.2
$R = (0, 1)$	$n_s = 500$	0.8850	0.4902	0.1364	0.0330	0.1358	0.4966	0.8708	0.9440
	$n_s = 750$	0.9908	0.8064	0.2806	0.0470	0.2420	0.7710	0.9870	0.9950
	$n_s = 1000$	1	0.9380	0.4372	0.0666	0.3618	0.9170	0.9998	1

Table IV reports the results on the null rejection probabilities for different sample sizes. The size is well controlled and is getting closer to 0.05 as the sample size increases for both hypotheses. Table V provides the finite sample power results when the model deviates from the null hypothesis. When the sample size is fixed, Table V shows that as the true value deviates further from the hypothesized value, the probability of rejecting the null hypothesis increases. At the same time, for any fixed deviation, the rejection probability increases with the sample size.

6.3 Separating Unobserved Heterogeneity from Multiple Equilibria

We use Games 3 and 4 to examine the performance of our estimator for the General game. For both games, we consider two designs: in Design 1, there are 2 equilibria on latent state A for the 8th observed state; and in Design 2, there are 2 equilibria on latent state A for both the 8th and the 16th observed state. For observed states on which there are no multiple equilibria, we let $\Pr(k = A \mid x) = 0.5$. For observed states on which there are multiple equilibria, the conditional distribution for the composite latent variable ω is specified as follows: $\omega = 1$ with probability 0.4, which corresponds to latent state A and equilibrium 1; $\omega = 2$ with probability 0.3, which corresponds to latent state A and equilibrium 2; $\omega = 3$ with probability 0.3, which corresponds to latent state B. The results of the simulations are reported in the table below. From 100 repetitions, we document the average running time and the correct grouping rate (CGR), where the running time is the total time to run our procedures developed in Sections 5.2.1 and 5.2.2, and the CGR is probability that $\widehat{\mathbf{S}}_{\widehat{K}}$ is equal to the correct partition. When $\widehat{\mathbf{S}}_{\widehat{K}}$ equals to the correct partition, $\widehat{K} = |\mathscr{K}|$ and $\widehat{t\pi}^k$ is the average

of consistent estimators of $t\pi_0^k$ for $k = 1, \ldots, |\mathscr{K}|$.

Table VI: Performance of the consistent grouping method on top of MMS: running time (in seconds) and correct grouping rate (CGR)

	Game 3				Game 4			
	Design 1		Design 2		Design 1		Design 2	
n_s	Time	CGR	Time	CGR	Time	CGR	Time	CGR
250	4.8851	0.99	5.0575	1	5.4538	0.96	5.9654	0.98
500	4.7580	1	5.0110	1	5.0065	0.98	5.0314	0.99
750	5.0105	1	4.6777	1	4.9731	1	5.0438	1
1000	4.7189	1	4.6639	1	4.8997	1	4.9798	1

From Table VI, we see that our estimation procedure developed in Section 5.2 for the General Game is both fast to run and accurate. The CGR is very close or equal to one across different designs and increases with sample size.

7 Conclusion

In this paper, we have proposed a computationally fast sequential method to estimate the payoff function and to conduct uniform inference in static games of incomplete information with unobserved heterogeneity and multiple equilibria. It builds on a novel characterization of the matching-types problem as a minimum-distance problem with both correct and incorrect moments. Based on this characterization, we develop a new MMS procedure that is extremely fast to implement. For inference, we construct an asymptotically uniformly valid test for linear hypotheses on the payoff function. The test is easy to implement with known critical values from the chi-squared distribution. An extensive Monte Carlo study is carried out to investigate the finite sample performance of our estimation and inference procedures.

Instead of employing sequential estimators, researchers could make use of the equilibrium condition to improve the finite sample performance of structural estimators. Aguirregabiria and Mira (2007) develop a nested pseudo-likelihood algorithm that imposes the equilibrium condition iteratively while avoids solving the equilibrium condition exactly for any parameter as in nested fixed point algorithm. In a companion paper we aim to extend the nested pseudo-likelihood algorithm to games with both multiple equilibria and unobserved heterogeneity.[25] We shall address open

[25] We thank an anonymous referee for suggesting this future line of research.

questions on the stability and convergence of the algorithm and to combine it with our consistent grouping method to separate multiple equilibria from unobserved heterogeneity.

The MMS procedure introduced in this paper has broad applicability besides the study of games. We are currently working on its extensions to the general moment selection problems discussed in Andrews (1999).

Appendix A Additional Details of the General Game

In this appendix, we provide additional details of the General Game discussed in Section 5. In Appendix A.1, we provide the expressions of $\overline{\pi}$ and Γ for constructing the system of moment functions. Appendix A.2 contains the detailed procedure for estimating $\left(c_0^h, \pi_0^h\right)$ for $h = 1, ..., |\Omega_{\mathbf{z}^1}|$ by extending the MMS procedure developed in Section 3.1 for the Simple Game. In Appendix A.3, we prove that our estimators are consistent and fast to compute. Appendix A.4 shows that the test is asymptotically uniformly valid and consistent.

A.1 Expressions of $\overline{\pi}$ and Γ

For $s = 1, \ldots, |\Omega_{\mathbf{z}}|$, we define $\omega(s, \mathbf{z})$ as the s-th value for the composite latent variable on observed state vector $\mathbf{z}$. The coefficient matrices in the moment function $G(\pi)$ for the General Game are

$$\overline{\pi} = \left[\overline{\pi}_1\left(\mathbf{z}^1\right)^\top, \ldots, \overline{\pi}_1\left(\mathbf{z}^l\right)^\top\right]^\top \text{ and } \Gamma = \left[\Gamma_1\left(\mathbf{z}^1\right)^\top, \ldots, \Gamma_1\left(\mathbf{z}^l\right)^\top\right]^\top,$$

where $\overline{\pi}$ has dimension $J\sum_{t=1}^{l}|\Omega_{\mathbf{z}^t}|$, Γ has dimension $J\sum_{t=1}^{l}|\Omega_{\mathbf{z}^t}|$ by $J(J+1)^{N-1}$, and for $t = 1, ..., l$,

$$\overline{\pi}_1\left(\mathbf{z}^t\right) = \begin{bmatrix} \overline{\pi}_1\left(1, \mathbf{z}^t, \omega\left(1, \mathbf{z}^t\right)\right) \\ \vdots \\ \overline{\pi}_1\left(J, \mathbf{z}^t, \omega\left(1, \mathbf{z}^t\right)\right) \\ \vdots \\ \overline{\pi}_1\left(1, \mathbf{z}^t, \omega\left(|\Omega_{\mathbf{z}^t}|, \mathbf{z}^t\right)\right) \\ \vdots \\ \overline{\pi}_1\left(J, \mathbf{z}^t, \omega\left(|\Omega_{\mathbf{z}^t}|, \mathbf{z}^t\right)\right) \end{bmatrix} \text{ and } \Gamma_1\left(\mathbf{z}^t\right) = \begin{bmatrix} \iota_1\left(1, \mathbf{z}^t\right) \\ \vdots \\ \iota_1\left(|\Omega_{\mathbf{z}^t}|, \mathbf{z}^t\right) \end{bmatrix}.$$

For $s = \{1, \ldots, |\Omega_{\mathbf{z}^t}|\}$, the matrix $\iota_1(s, \mathbf{z}^t)$ is a block diagonal matrix with J identical blocks given by:

$$\iota_1\left(s, \mathbf{z}^t\right) = \begin{bmatrix} \mathbf{p}_{-1}\left(\mathbf{z}^t, \omega\left(s, \mathbf{z}^t\right)\right) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{p}_{-1}\left(\mathbf{z}^t, \omega\left(s, \mathbf{z}^t\right)\right) \end{bmatrix},$$

where $\mathbf{p}_{-1}\left(\mathbf{z}^t, \omega\left(s, \mathbf{z}^t\right)\right)$ is a row vector with $(J+1)^{N-1}$ elements being the probabilities of joint actions for all other players in the same spirit as the Simple Game in Section 2.2.2.

A.2 MMS Procedure for Estimating π_0^h

We focus on developing the MMS procedure for (c_0^1, π_0^1) to simplify our discussion and notation.

Let sc denote the sub-selection vector of dimension $J\sum_{t=1}^{l}|\Omega_{\mathbf{z}^t}|$ that consists of $\mathbf{e}_1$ and $\mathbf{e}_0$. Following Definition 5.1, denote sc_t for $t = 1, ..., l$ as the t-th subvector of sc such that $sc \equiv [sc_1, \ldots, sc_l]^\top$. Define $J_n(sc) \equiv \min_{\pi \in \Pi} \|G_{n,sc}(\pi)\|^2$.

Step 0: Set $l_1 \in \{l_\pi, l_\pi + 1, .., l\}$, $\alpha_1 \in (0, 1]$, $\lambda \in (-1, 0)$, and $\Delta \in \{1, .., l - l_1\}$. Let $S = \left\lceil \frac{l - l_1}{\Delta} \right\rceil$ and $\alpha = \left(2^{|\Omega_{\mathbf{z}^1}|} - 1\right)^{-\Delta}$. The value of α is chosen according to the same spirit as the one in the MMS procedure for the Simple Game: more moments we add in each step, smaller proportion of matchings we tend to keep. See Section 3.3 for more detail. Relabel $\left\{\mathbf{z}^1, \ldots, \mathbf{z}^l\right\}$ such that $|\Omega_{\mathbf{z}}|$ for $\mathbf{z} \in \left\{\mathbf{z}^1, \ldots, \mathbf{z}^l\right\}$ are in an ascending order.

Step 1: The input of Step 1 is

$$\mathscr{SC}^1 \equiv \left\{ \begin{array}{c} [sc_1, ..., sc_l]^\top : sc_1 = \left[sc_{1,1}, \ldots, sc_{1,|\Omega_{\mathbf{z}^1}|}\right] \text{ with } sc_{1,1} = \mathbf{e}_1 \text{ and } sc_{1,j} \in \{\mathbf{e}_1, \mathbf{e}_0\} \text{ for } j \neq 1; \\ \text{for } t = 2, \ldots, l_1,\ sc_t = \left[sc_{t,1}, \ldots, sc_{t,|\Omega_{\mathbf{z}^t}|}\right] \text{ with } sc_{t,\mathrm{w}} \in \{\mathbf{e}_1, \mathbf{e}_0\}, \text{ where} \\ \mathrm{w} \in \{1, \ldots, |\Omega_{\mathbf{z}^t}|\} \text{ and } sc_t \neq [\mathbf{e}_0, \ldots, \mathbf{e}_0];\ \text{for } t = l_1 + 1, \ldots, l,\ sc_t = [\mathbf{e}_0, \ldots, \mathbf{e}_0] \end{array} \right\}.$$

Sort $J_n(sc^1)$ for all $sc^1 \in \mathscr{SC}^1$, and denote $J_n^{\alpha_1}$ as the value of the $100\alpha_1\%$ smallest. The output of Step 1 is

$$\mathscr{SC}_n^1 \equiv \left\{sc^1 \in \mathscr{SC}^1 : J_n\left(sc^1\right) \leq \max\left\{J_n^{\alpha_1}, n^\lambda\right\}\right\}.$$

Steps 2, 3, ... S: For $s = 2, ..., S$, define $l_s \equiv l_{s-1} + \Delta$. The input of Step s is the collection of sub-selection vectors defined as

$$\mathscr{SC}^s \equiv \left\{ \begin{array}{c} [sc_1, \ldots, sc_l]^\top : \left[sc_1, \ldots, sc_{l_{s-1}}\right] = \left[sc_1^{s-1}, \ldots, sc_{l_{s-1}}^{s-1}\right] \text{ for some } sc^{s-1} \in \mathscr{SC}_n^{s-1}; \\ \text{for } t = l_{s-1} + 1, \ldots, l_s,\ sc_t = \left[sc_{t,1}, \ldots, sc_{t,|\Omega_{\mathbf{z}^t}|}\right] \text{ with } sc_{t,\mathrm{w}} \in \{\mathbf{e}_1, \mathbf{e}_0\}, \text{ where} \\ \mathrm{w} \in \{1, \ldots, |\Omega_{\mathbf{z}^t}|\} \text{ and } sc_t \neq [\mathbf{e}_0, \ldots, \mathbf{e}_0];\ \text{for } t = l_s + 1, \ldots, l,\ sc_t = [\mathbf{e}_0, \ldots, \mathbf{e}_0] \end{array} \right\}.$$

The output of Step s is

$$\mathscr{SC}_n^s \equiv \left\{ sc^s \in \mathscr{SC}^s : J_n\left(sc^s\right) \leq \max\left\{ J_n^\alpha, n^\lambda \right\} \right\}.$$

Step $(\mathbf{S+1})$: The *effective parameter space* for c_0^1 is

$$\mathscr{C}_n^1 \equiv \left\{ \begin{array}{c} [c_1, \ldots, c_l]^\top : [c_1, \ldots, c_{l_S}] = \left[sc_1^S, \ldots, sc_{l_S}^S\right] \text{ for some } sc^S \in \mathscr{SC}_n^S; \\ \text{for } t = l_S + 1, \ldots, l,\ c_t = \left[c_{t,1}, \ldots, c_{t,|\Omega_{\mathbf{z}^t}|}\right] \text{ with } c_{t,\mathrm{w}} \in \{\mathbf{e}_1, \mathbf{e}_0\}, \\ \text{where } \mathrm{w} \in \{1, \ldots, |\Omega_{\mathbf{z}^t}|\} \text{ and } c_t \neq [\mathbf{e}_0, \ldots, \mathbf{e}_0] \end{array} \right\}.$$

Given the effective parameter space, the MMS estimator for (c_0^1, π_0^1) is defined as

$$\left(\widehat{c}^1, \widehat{\pi}^1\right) \equiv \arg \min_{c \in \mathscr{C}_n^1, \pi \in \Pi} \left\| G_{n,c}\left(\pi\right) \right\|_{W_n(c)}^2 - \rho_1\left(\|c\|_0\right) \kappa_{1,n}/n.$$

By eliminating incorrect matchings in multiple steps, the size of the effective parameter space $\mathscr{C}_n^1$ is much smaller than that of $\mathscr{C}^1$. As a result, the multistep estimator $(\widehat{c}^1, \widehat{\pi}^1)$ is much faster to compute than $(\widetilde{c}^1, \widetilde{\pi}^1)$. We repeat the above procedure to obtain estimators $\left(\widehat{c}^h, \widehat{\pi}^h\right)$ for $h = 2, ..., |\Omega_{\mathbf{z}^1}|$.

A.3 Asymptotic Properties of the Estimators

In this section we present results on the consistency of our estimators and the time and space complexities of our MMS procedure. We provide sufficient conditions that parallel those stated in the Simple Game. The proofs for the results in this section follow the similar arguments for the results in Sections 3.2. All proofs are collected in the online appendix.

We first provide assumptions for the consistency of the proposed estimator $\left(\widehat{c}^h, \widehat{\pi}^h\right)$ for $h = 1, ..., |\Omega_{\mathbf{z}^1}|$.

Assumption A.1. *(i)* Π *is compact. (ii) The estimated equilibrium CCPs are* $\sqrt{n}$ *consistent and asymptotically normal.*

Assumption A.1 (ii) is satisfied by many estimators such as the ones proposed by Bonhomme et al. (2016) and Xiao (2018). See Appendix C.3 for primary conditions that guarantee the consistency and asymptotic normality of Xiao (2018)'s CCP estimator.

Assumption A.2. $\forall\ c \in \mathscr{C}_n^h$ *for* $h = 1, ..., |\Omega_{\mathbf{z}^1}|$, $W_n\left(c\right) \xrightarrow{p} W\left(c\right)$ *for some positive definite matrix* $W\left(c\right)$.

Assumption A.3. *$\rho_1(\cdot) > 0$ is a known strictly increasing function and $\kappa_{1,n} \to \infty$ with $\kappa_{1,n} = o(n)$.*

Theorem A.1. *Under Assumptions 5.1-5.3 and A.1-A.3, it holds that for $h = 1, ..., |\Omega_{\mathbf{z}^1}|$, $\widehat{c}^h = c_0^h$ wp $\to 1$ and $\widehat{\pi}^h \xrightarrow{p} \pi_0^h$ for any $l_1 \in \{l_\pi, l_\pi + 1, .., l\}$, $\alpha_1 \in (0, 1]$, $\lambda \in (-1, 0)$, and $\Delta \in \{1, .., l - l_1\}$.*

The time and space complexities of the MMS procedure achieve linearity in l if the following assumption holds.

Assumption A.4. *The number of the observed states where multiple equilibria exist is not a function of l.*

The MMS procedure improves the computational time upon the classical moment selection estimator by eliminating selection vectors $c \in \mathscr{C}$ such that $c \notin \mathscr{C}\mathscr{I}$ in multiple steps. In the Simple Game, $\mathscr{C}\mathscr{I}$ contains only one element; while in the General Game, $\mathscr{C}\mathscr{I}$ can contain more than one element because of multiple equilibria. In cases where multiple equilibria occur at a significant portion of the observed states, the number of elements in $\mathscr{C}\mathscr{I}$ is large and might be of similar magnitude as $\mathscr{C}$. The MMS procedure will no longer enjoy the linear time complexity because very few selection vectors can be eliminated in such cases. However, even when the number of observed states is large, it is not uncommon to have multiple equilibria being played at only a few observed states.

Theorem A.2. *Let Assumptions 5.1-5.3 and A.1-A.4 hold. Then with probability approaching one as $n \to \infty$, for all payoffs except for a set of Lebesgue measure zero, both the time and space complexities of the MMS procedure for the General Game are linear in l.*[26]

With the following assumption on $\rho_2(\cdot)$ and $\kappa_{2,n}$, we can show that $\widehat{K}$ and $\widehat{t\pi}^k$ for $k = 1, \ldots, \widehat{K}$ are consistent estimators.

Assumption A.5. *$\rho_2(\cdot) > 0$ is a known strictly increasing function and $\kappa_{2,n} \to \infty$ with $\kappa_{2,n} = o(n)$.*

Theorem A.3. *Under Assumptions 5.1-5.3, A.1-A.3, and A.5, $\widehat{K} = |\mathscr{K}|$ wp $\to 1$; and $\widehat{t\pi}^k$ with $k = 1, \ldots, \widehat{K}$ are consistent estimators for $t\pi_0^k$ with $k = 1, \ldots, |\mathscr{K}|$.*

[26] In Lemma B.9 of the online appendix, we show that the time complexity of the MMS procedure is at most a polynomial function in l as long as the number of observed states where multiple equilibria exist grows slowly with l.

A.4 Asymptotic Validity and Consistency of the Inference Procedure

For controlling the asymptotic size, we impose the following assumptions. The definitions of Ξ_R and $\Xi_R(\xi)$ extend the ones in Section 4.

Assumption A.6. *(i) On any* $\mathbf{z}$ *and* ω, $\frac{\partial^2 \Lambda_{1\mathbf{z}\omega}^{-1}(\mathbf{p})}{\partial \mathbf{p} \partial \mathbf{p}^\top}$ *is bounded for any* $\mathbf{p}$. *(ii) For any* $\xi \in \Xi_R$, $\mathbf{z}$ *takes each value in* $\mathscr{Z}$ *with probability bounded below by* $\varepsilon > 0$. *(iii) For any* $\xi \in \Xi_R$ *and the parameter sequence* $\{\xi_n\} \in \Xi_R(\xi)$, *given each* $c \in \mathscr{C}_n^h$, $W_n(c) = W(c) + o_p(1)$, *with* $W(c)$ *being positive definite. (iv)* $W\left(c_0^h\right) = \Omega_{h,0}^{-1}$ *for* $\Omega_{h,0}$ *being the asymptotic variance of* $\sqrt{n}\left(G_{n,c_0^h}\left(\pi_0^h\right) - G_{c_0^h}\left(\pi_0^h\right)\right)$.

Assumption A.6 (i) and (ii) play the same role as Assumption 4.1 (i) and (ii), and they are needed for uniform asymptotic linear representation of the moment function. In particular, Assumption A.6 (i) is satisfied, for example, if the normalized private information follows a logistic distribution. Assumption A.6 (iii) and (iv) are standard assumptions on the weighting matrix.

The following theorem shows that using an appropriately chosen chi-squared critical value, the test achieves asymptotic size control.

Theorem A.4. *Let Assumptions 5.1-5.3, A.1, and A.6 hold. For any* $l_1 \in \{l_\pi, l_\pi + 1, .., l\}$, $\alpha_1 \in (0, 1]$, $\lambda \in (-1, 0)$, *and* $\Delta \in \{1, .., l - l_1\}$, *it holds that*

$$\limsup_{n\to\infty} \sup_{\xi \in \Xi_R} \Pr_\xi\left(T_n > \chi^2_{[Jl - l_\pi + l_R], 1-\alpha}\right) \leq \alpha.$$

In Theorem 4.1, the subscript l in the critical value corresponds to the number of moments selected by elements in $\mathscr{C}$. In the General Game, elements in $\mathscr{C}^h$ select at least Jl number of moments. In addition, there might be more than one element in $\mathscr{C}\mathscr{I}^h$. We obtain inequality in Theorem A.4 rather than equality in Theorem 4.1.

To obtain consistency of the test, we need to impose the following assumption.

Assumption A.7. $G_c(\pi) = \mathbf{0}$ *has a unique solution for any* $c \in \mathscr{C}\mathscr{I}^h$.

Without Assumption A.7, there might exist some $c^\star \in \mathscr{C}\mathscr{I}^h$ and $\pi^\star \neq \pi_0^h$ such that $G_{c^\star}(\pi^\star) = \mathbf{0}$. The asymptotic power of the test is not one if $R\pi^\star = r$. In the Simple Game where there is no multiple equilibria, $\mathscr{C}\mathscr{I}^h$ contains only one element c_0^h. By the identification assumption, $G_{c_0^h}(\pi) = \mathbf{0}$ has a unique solution, making Assumption A.7 automatically satisfied. However, the identification assumption for

the General Game, Assumption 5.3, allows $G_c(\pi) = \mathbf{0}$ to have multiple solutions for $c \in \mathscr{C}\mathscr{I}^h$ and $c \neq c_0^h$. Assumption A.7 is essential to guarantee the consistency. With the same argument in the proof of Theorem 3.2, it can be shown that Assumption A.7 holds for all payoffs except for a set of Lebesgue measure zero.

It is also worth mentioning that to achieve both the asymptotic size control and consistency, it is difficult if not impossible to avoid imposing Assumption A.7 by modifying the test statistic. To obtain the consistency without Assumption A.7, we need to find a criterion that separates c_0^h with other $c \in \mathscr{C}\mathscr{I}^h$. Since they barely differ by their $\|\cdot\|_0$ norms, the norm is the only criterion. However, under drifting model parameter sequences, restricting ourselves to selection vectors in $\mathscr{C}^h$ that have the largest $\|\cdot\|_0$ norm may result in excluding c_0^h due to the existence of nearly true selections. For a nearly true selection $c^\dagger \in \mathscr{C}^h$ such that $G_{c^\dagger}^{\xi_n}(\pi) \neq \mathbf{0}$ but $G_{c^\dagger}^{\xi_n}(\pi) \to \mathbf{0}$ for some $\pi \neq \pi_0^h$ as $n \to \infty$, it is possible that $\left\|c^\dagger\right\|_0 > c_0^h$. In consequence, Assumption A.7 is crucial for the consistency of any test.

Theorem A.5. *Let Assumptions 5.1-5.3, A.1, A.2 and A.7 hold. For any $l_1 \in \{l_\pi, l_\pi + 1, .., l\}$, $\alpha_1 \in (0, 1]$, $\lambda \in (-1, 0)$, and $\Delta \in \{1, .., l - l_1\}$, it holds that for any $\xi \notin \Xi_R$*

$$\lim_{n\to\infty} \Pr_\xi\left(T_n > \chi^2_{[Jl - l_\pi + l_R], 1-\alpha}\right) = 1.$$

For implementation of the test using the bootstrap weighting matrix $W_n^b(c)$, we need to match the labels of composite latent variables across different bootstrap draws. The following assumption provides a sufficient condition. Define

$$\begin{aligned}
&\mathbf{P}(\mathbf{z}, \omega) \equiv [\mathbf{p}_1(\mathbf{z}, \omega), \ldots, \mathbf{p}_N(\mathbf{z}, \omega)], \text{ where} \\
&\mathbf{p}_i(\mathbf{z}, \omega) \equiv [\Pr(d_i = 1 \mid \mathbf{z}, \omega), \ldots, \Pr(d_i = J \mid \mathbf{z}, \omega)].
\end{aligned}$$

Assumption A.8. *There is a known scalar valued function $\psi(\cdot)$ such that for each $\mathbf{z} \in \mathscr{Z}$ and $\omega \neq \omega'$, $\psi\Big(\mathbf{P}(\mathbf{z}, \omega), p(\omega \mid \mathbf{z})\Big) \neq \psi\Big(\mathbf{P}(\mathbf{z}, \omega'), p(\omega' \mid \mathbf{z})\Big)$.*

Examples $\psi(\cdot)$ include $\psi(\mathbf{P}(\mathbf{z}, \omega), p(\omega \mid \mathbf{z})) = p(\omega \mid \mathbf{z})$ and $\psi\left(\mathbf{P}(\mathbf{z}, \omega), p^k(\mathbf{z})\right) = \sum_{i=1}^N \Pr(d_i = 1 \mid \mathbf{z}, \omega)$. Researchers can track ω using the distinctive weight corresponding to each ω in the first case or using the sum of equilibrium CCPs of choosing 1 across all players in the second case. Note that Assumption A.8 is not needed for the Simple Game because of Assumption 2.5. Under Assumptions 5.1-5.3, A.1, A.6 (i) and (ii), and A.8, the bootstrap weighting matrix satisfies Assumption A.6 (iii) and (iv).

Acknowledgment: This paper is a revised version of the previously circulated paper "Estimation and Inference in Games of Incomplete Information with Non-separable Unobserved Heterogeneity," dated September 17, 2020. We would like to thank the Editor Stéphane Bonhomme, a senior member of the editorial board, and three anonymous referees for their insightful comments, which substantially improved this paper. We thank Aureo de Paula, Jin-Chuan Duan, Arnaud Maurel, Xun Tang, and Ruli Xiao, and participants of the 2018 Seattle-Vancouver Econometrics Conference, the 2019 North American Summer Econometrics Society Meetings, the 2021 Annual Conference of the International Association for Applied Econometrics, and 2023 Econometric Society Australasian Meeting for helpful discussions. This work was facilitated by Hyak supercomputer system and CSDE simulation cluster at the University of Washington. Jiang acknowledges the financial support from the Humanities and Social Sciences Foundation of the Ministry of Education of China Grant [grant number 23YJC790053], Basic Scientific Center Project [grant number 71988101] of National Science Foundation of China, the 111 Project [grant number B13028] and National Natural Science Foundation of China [grant number 72373127].

References

Aguirregabiria, V. and M. Marcoux (2021): "Imposing Eequilibrium Restrictions in the Estimation of Dynamic Discrete Games," *Quantitative Economics*, 12, 1223–1271.

Aguirregabiria, V. and P. Mira (2007): "Sequential Eestimation of Dynamic Discrete Games," *Econometrica*, 75, 1–53.

——— (2019): "Identification of Games of Incomplete Information with Multiple Equilibria and Common Unobserved Heterogeneity," *Quantitative Economics*, 10, 1659–1701.

Allman, E. S., C. Matias, and J. A. Rhodes (2009): "Identifiability of Parameters in Latent Structure Models with Many Observed Variables," *The Annals of Statistics*, 37, 3099–3132.

Andrews, D. W. (1999): "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67, 543–563.

Andrews, D. W. and X. Cheng (2012): "Estimation and Inference with Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211.

Andrews, D. W., X. Cheng, and P. Guggenberger (2020): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," *Journal of Econometrics*.

Andrews, D. W. and B. Lu (2001): "Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models," *Journal of Econometrics*, 101, 123–164.

Aradillas-Lopez, A. (2010): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," *Journal of Econometrics*, 157, 409–431.

Aradillas-López, A., A. Gandhi, and D. Quint (2016): "A Simple Test for Moment Inequality Models with an Application to English Auctions," *Journal of Econometrics*, 194, 96–115.

Bajari, P., C. L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75, 1331–1370.

Bajari, P., H. Hong, and D. Nekipelov (2013): "Game Theory and Econometrics: A Survey of Some Recent Research," in *Advances in economics and econometrics, 10th world congress*, vol. 3, 3–52.

Bajari, P., H. Hong, and S. P. Ryan (2010): "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica*, 78, 1529–1568.

Bergemann, D. and S. Morris (2013): "Robust Predictions in Games with Incomplete Information," *Econometrica*, 81, 1251–1308.

——— (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522.

Bonhomme, S., K. Jochmans, and J.-M. Robin (2016): "Estimating Multivariate Latent-Structure Models," *The Annals of Statistics*, 44, 540–563.

Bonhomme, S., T. Lamadon, and E. Manresa (2022): "Discretizing Unobserved Heterogeneity," *Econometrica*, 90, 625–643.

Cheng, X. (2015): "Robust Inference in Nonlinear Models with Mixed Identification Strength," *Journal of Econometrics*, 189, 207–228.

Cheng, X. and Z. Liao (2015): "Select the Valid and Relevant Moments: An Information-based LASSO for GMM with Many Moments," *Journal of Econometrics*, 186, 443–464.

De Paula, A. (2013): "Econometric Analysis of Games with Multiple Equilibria," *Annual Review of Economics*, 5, 107–131.

De Paula, A. and X. Tang (2012): "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," *Econometrica*, 80, 143–172.

Grieco, P. L. (2014): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets," *The RAND Journal of Economics*, 45, 303–340.

Haile, P. A. and E. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1–51.

Hotz, V. J. and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60, 497–529.

Igami, M. and N. Yang (2016): "Unobserved Heterogeneity in Dynamic Games: Cannibalization and Preemptive Entry of Hamburger Chains in Canada," *Quantitative Economics*, 7, 483–521.

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

Leeb, H. and B. M. Pötscher (2005): "Model Selection and Inference: Facts and Fiction," *Econometric Theory*, 21, 21–59.

——— (2008): "Can One Estimate the Unconditional Distribution of Post-Model-Selection Estimators?" *Econometric Theory*, 24, 338–376.

Lewbel, A. and X. Tang (2015): "Identification and Estimation of Games with Incomplete Information using Excluded Regressors," *Journal of Econometrics*, 189, 229–244.

Li, S., Y. Liu, and K. Deininger (2013): "How Important are Endogenous Peer Effects in Group Lending? Estimating a Static Game of Incomplete Information," *Journal of Applied Econometrics*, 28, 864–882.

Liao, Z. (2013): "Adaptive GMM Shrinkage Estimation with Consistent Moment Selection," *Econometric Theory*, 29, 857–904.

Liu, N., Q. Vuong, and H. Xu (2017): "Rationalization and Identification of Discrete Games with Correlated Types," *Journal of Econometrics*, 201, 249–268.

Lloyd, S. (1982): "Least Squares Quantization in PCM," *IEEE transactions on information theory*, 28, 129–137.

Luo, Y., P. Xiao, and R. Xiao (2022): "Identification of Dynamic Games with Unobserved Heterogeneity and Multiple Equilibria," *Journal of Econometrics*, 226, 343–367.

Magnolfi, L. and C. Roncoroni (2023): "Estimation of Discrete Games with Weak Assumptions on Information," *The Review of Economic Studies*, 90, 2006–2041.

Pesendorfer, M. and P. Schmidt-Dengler (2008): "Asymptotic least squares estimators for dynamic games," *The Review of Economic Studies*, 75, 901–928.

Sweeting, A. (2009): "The Strategic Timing Incentives of Commercial Radio Stations: An Empirical Analysis Using Multiple Equilibria," *The RAND Journal of Economics*, 40, 710–742.

Wan, Y. and H. Xu (2014): "Semiparametric Estimation of Binary Decision Games of Incomplete Information with Correlated Private Signals," *Journal of Econometrics*, 182, 235–246.

Xiao, R. (2018): "Identification and Estimation of Incomplete Information Games with Multiple Equilibria," *Journal of Econometrics*, 203, 328–343.

Xu, H. (2014): "Estimation of Discrete Games with Correlated Types," *The Econometrics Journal*, 17, 241–270.

Zhu, T. and V. Singh (2009): "Spatial Competition with Endogenous Location Choices: An Application to Discount Retailing," *Quantitative Marketing and Economics*, 7, 1–35.