Identification of Games of Incomplete Information with Multiple Equilibria and Unobserved Heterogeneity

Victor Aguirregabiria*
University of Toronto and CEPR

Pedro Mira*
CEMFI, Madrid

This version: January 26th, 2019

Abstract

This paper deals with identification of discrete games of incomplete information when we allow for three types of unobservables: payoff-relevant variables, both players' private information and common knowledge; and non-payoff-relevant variables that determine the selection between multiple equilibria. The specification of the payoff function and the distributions of the common knowledge unobservables is nonparametric with finite support (i.e., finite mixture model). We provide necessary and sufficient conditions for the identification of all the primitives of the model. Two types of conditions play a key role in our identification results: independence between players' private information, and an exclusion restriction in the payoff function. When using a sequential identification approach, we find that the up-to-label-swapping identification of the finite mixture model in the first step creates a problem in the identification of the payoff function in the second step: unobserved types have to be correctly matched across different values of observable explanatory variables. We show that this matching-types problem appears in the sequential estimation of other structural models with nonparametric finite mixtures. We derive necessary and sufficient conditions for identification, and show that additive separability of unobserved heterogeneity in the payoff function is a sufficient condition to deal with this problem. We also compare sequential and joint identification approaches.

Keywords: Discrete games of incomplete information; Multiple equilibria in the data; Unobserved heterogeneity; Finite mixture models; identification up to label swapping.

JEL codes: C13, C35, C57.

*A previous version of this paper was titled *"Structural Estimation in Games when the Data Come fromMultiple Equilibria"*. We would like to thank comments from the Editor, three anonymous referees, Juan Esteban, Yingyao Hu, Nicolai Kuminoff, Sokbae Lee, Mathieu Marcoux, John Rust, Steven Stern, Artyom Shneyerov, Xun Tan, Ken Wolpin, and from participants in seminars at Concordia, Copenhagen, Duke University, Texas A&M, University College of London, the Econometric Society World Congress, and the UPenn conference in honor of Kenneth I. Wolpin. This research was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) and by the Ministerio de Ciencia e Innovación (MICINN), grant ECO2008-00280.

1 Introduction

Multiplicity of equilibria is a prevalent feature in games. An implication of multiplicity of equilibria in the structural estimation of games is that the model predicts more than one probability distribution of the endogenous variables. The standard criteria used for estimation, such as likelihood or GMM criteria, are no longer functions of the structural parameters but correspondences, and this makes the application of these estimation methods impractical in many relevant cases. A substantial part of the recent literature on the econometrics of games of incomplete information proposes simple two-step estimators that deal with these issues.[1] These two-step methods assume that there are no unobservables that are common knowledge to players, and that the same equilibrium has been played in all the observations in the data. The model may have multiple equilibria for the true value of the structural parameters, but only one of them is present in the data.[2] Under these assumptions, structural parameters in these models are identified given the same type of exclusion restrictions as in games with equilibrium uniqueness (see Bajari et al., 2010).

The assumption that all the data have been generated from a single equilibrium is very strong. In most empirical games of incomplete information in the literature, uniqueness of the equilibrium in the data, together with the assumption that there are no common knowledge unobservables, imply that the actions of players are independent of one another conditional on observables. This testable implication is likely to fail in most datasets. One possible interpretation of failure of this conditional independence is that common knowledge unobservables are present.[3] An alternative interpretation is multiple equilibria in the data.[4] These two alternative explanations can generate different estimations of the structural parameters and different predictions when we use the estimated model to make counterfactual experiments. Therefore, a relevant question is whether it is

[1] See Aguirregabiria and Mira (2007), Bajari et al. (2007), and Pesendorfer and Schmidt-Dengler (2008) as seminal contributions in this literature. Other recent contributions to this topic in the context of games of incomplete information are Sweeting (2009), Aradillas-Lopez (2010), and Bajari, Hong, Krainer, and Nekipelov (2010). See Bajari, Hong, and Nekipelov (2013) for a survey of this literature.

[2] A weaker version of this assumption establishes that we can partition the data into a number of subsamples according to the value of an exogenous variable such that the same equilibrium is played within each subsample.

[3] Aguirregabiria and Mira (2007), Arcidiacono and Miller (2011) extend sequential estimation methods to allow for common knowledge unobservables in games of incomplete information. They do not allow for multiple equilibria in the data and consider parametric models.

[4] De Paula and Tang (2012) relax the assumption of a unique equilibrium in the data. They interpret failure of independence in terms of multiple equilibria and show that it is actually helpful to identify the sign of the parameters that capture the strategic interactions between players. However, de Paula and Tang assume that the model does not contain common knowledge unobservables.

possible to identify from the data the contribution of unobservables that affect the selection of an equilibrium from the contribution of unobservables that are payoff-relevant.

Authors in different areas of economics have proposed multiplicity of equilibria as a plausible explanation for important economic phenomena. This argument has been used in empirical applications to explain bank runs (Cooper and Corbae, 2002, and Egan, Hortacsu, and Matvos, 2017), spatial distribution of economic activity (Krugman, 1991, Davis and Weinstein, 2002, 2008, and Bayer and Timmins, 2005, 2007), macroeconomic fluctuations (Farmer and Guo, 1995), market variation in firms' behavior (Sweeting, 2009, Ellickson and Misra, 2008, and Grieco, 2014), and changes in wage inequality (Moro, 2003), among others. In all these applications, the identification of the contribution of multiple equilibria has been based on strong restrictions on the role of payoff-relevant unobserved heterogeneity, e.g., ruling out this form of heterogeneity. One of the main purposes of this paper is to obtain conditions for the identification of the relative contribution of multiple equilibria and payoff-relevant unobservables when both sources of unobserved heterogeneity are specified nonparametrically and allowed to have the same degree of variation.

In this paper, we study the identification of games when we allow for three types of unobserved heterogeneity for the researcher: payoff-relevant, player-specific variables or "types" that are private information (PI unobservables); payoff-relevant variables that are common knowledge to all the players (PR unobservables); and variables that are common knowledge to all the players and are not payoff-relevant but affect the equilibrium selection (multiple equilibria or ME unobservables). The specification of the payoff function is nonparametric, and the probability distribution of common knowledge unobservables is also nonparametric but with finite support (i.e., finite mixture model). The model is semiparametric because we assume that the researcher knows the distribution of the private information unobservables, which are independent across players, up to a scale parameter.

As far as we know, this is the first paper to study nonparametric identification of games with these three different sources of unobservables. More specifically, the model in this paper extends the specifications of several important papers in the literature on identification of games. Sweeting (2009) and De Paula and Tang (2012) allow for multiple equilibria but not for PR unobservables. Otsu, Pesendorfer and Takahashi (2016) consider games of incomplete information with common knowledge unobserved heterogeneity that can be either PR or ME. The paper proposes a test for the existence of unobserved heterogeneity using panel data but it does not deal with the identifi-

cation of payoffs or with the separate identification of the contribution of PR and ME unobserved heterogeneity. Our model is similar to the one in Grieco (2014). Grieco considers a game of market entry that includes the three types of unobservables in our model, i.e., PI, PR, and ME unobservables. Grieco's model is fully parametric in the specification of the payoff function, the distribution of the PR unobservables, and the distribution of the equilibrium selection. The identification results in Grieco's paper rely crucially on these parametric restrictions. In this paper, we consider identification conditions that are not based on parametric assumptions.

We show that, in a model with N players, $J+1$ choice alternatives, L points of support in the distribution of common knowledge unobservables, and with $N \geq 3$ and $L \leq (J+1)^{(N-1)/2}$, all the structural functions of the model are identified under the same type of exclusion restrictions that we need for identification without unobserved heterogeneity. In particular, we can separately identify the relative contributions of payoff-relevant and multiple equilibria unobserved heterogeneity to explain players' behavior.

Two types of conditions play a key role in our identification results: independence between players' private information, and an exclusion restriction in the payoff function. Most of our identification results in this paper are based on a sequential approach. In a first step, we consider the nonparametric identification of players' strategies (defined as Conditional Choice Probabilities) and the distribution of common knowledge unobservables in the context of a nonparametric finite mixture model. The key identifying restriction in this first step is the independence between players' private information variables. In a second step, we study the identification of payoffs and the separate identification of payoff-relevant (PR) and multiple-equilibria (ME) common knowledge unobservables. Identification in this second step is based on a exclusion restriction on players' payoff functions. We show that the conditions for the identification of the finite mixture model in the first step are sufficient but not necessary. In particular, when using a non-sequential identification approach, the exclusion restrictions in the payoff function can help us to relax some of the restrictions that we use to identify the finite mixture model in the first step of the sequential approach.

We also find an important and previously neglected issue in the implementation of the sequential identification approach. In the identification of the finite mixture model in the first step, it is well known that the distribution of the unobservables is identified up to label swapping of the types. We can identify the distribution of the unobservables for each value of the exogenous variables but,

without further assumptions, we cannot "match" unobservable types across different values of these exogenous variables. We show that this up-to-label-swapping identification in the first step creates a problem in the identification of the payoff function in the second step: unobserved types have to be correctly matched across different values of observable explanatory variables. We also show that this matching-types problem appears in the sequential estimation of other structural models with nonparametric finite mixtures, such as single-agent models, static or dynamic. We derive necessary and sufficient conditions for identification under this problem, and show that additive separability of unobserved heterogeneity in the payoff function is a sufficient condition to deal with this problem. We also present and discuss the relative merits and limitations of other sufficient conditions for identification such as independence between unobservables and explanatory variables.

Our identification results and tests rely on the assumption that players' actions in the data come from Bayesian Nash equilibria (BNE) for a particular information structure, one where information is either common-knowledge to all the players or privately known by only one player, and this player-specific private information is unobservable to the researcher. While our framework generalizes the specification of the unobservables in the most widely-used class of empirical games of incomplete information, the assumption on the information structure is maintained. One might consider environments with the same basic specification as ours but different informational assumptions. For instance, some information might be shared by a subgroup of players. Or players might receive signals containing information about the other players' types. In recent work, Bergemann and Morris (2013, 2016) have introduced the Bayesian Correlated Equilibrium (BCE) as a solution concept which is more robust, in the sense that it delivers all predictions compatible with Bayesian Nash equilibria for any information structure within a wide class. Magnolfi and Roncoroni (2017) study inference based on the BCE solution concept. Their goal is to identify only the payoff parameters, and their work illustrates a tradeoff between robustness to assumptions about information structures and the ability to achieve point identification.

The rest of the paper is organized as follows. Section 2 introduces the class of models. Section 3 describes the type of data and the assumptions on the data generating process. Section 4 presents our identification results using a sequential approach. Section 5 presents necessary and sufficient conditions for local joint identification and a numerical example of a class of 2-player games which are jointly identified but not sequentially identified locally. We conclude in section 6.

2 Model

Consider a game that is played by N players which are indexed by $i \in \mathcal{I} = \{1, 2, ..., N\}$. Each player has to choose an action from a discrete set of alternatives $\mathcal{A} = \{0, 1, ..., J\}$. The decision of player i is represented by the variable $a_i \in \mathcal{A}$. Each player chooses his action a_i to maximize his expected payoff. The payoff function of player i is $\Pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \omega, \boldsymbol{\varepsilon}_i)$, where: $\Pi_i(.)$ is a real-valued function; $\boldsymbol{a}_{-i} \in \mathcal{A}^{N-1}$ is a vector with choice variables of players other than i; and $\mathbf{x} \in \mathcal{X}$, $\omega \in \Omega$, and $\boldsymbol{\varepsilon}_i$ are vectors of exogenous characteristics of players and of the environment (market). The variables in $\mathbf{x}$ and ω are common knowledge for all players, and the vector $\boldsymbol{\varepsilon}_i$ is private information of player i. Variables ω and $\boldsymbol{\varepsilon}_i$ are unobservable to the researcher and $\mathbf{x}$ is observable.

In addition to these payoff relevant state variables, there are also common knowledge, non-payoff relevant variables that affect players' beliefs about which equilibrium, from the multiple ones the model has, is the one that they are playing. We denote these as *sunspots* and represent them using the vector ξ. These *sunspot* variables are unobservable to the researcher. For the rest of the paper, we denote the unobservables $\boldsymbol{\varepsilon}_i$ as PI (for private information), ω as PR (for payoff relevant), and ξ as ME (for multiple equilibria).

EXAMPLE 1: Coordination game within the classroom (Todd and Wolpin, 2012). In a school class the students and the teacher choose their respective levels of effort, $a_i \in \mathcal{A}$. Each student has preferences on her own end-of-the-year knowledge, Π_i. The teacher cares about the aggregate knowledge of all the students. A student's knowledge depends on her own effort, the effort of her peers, teacher's effort, and exogenous characteristics of the student, the classroom, and the school. This type of game is an example of *Coordination Game* (Cooper, 1999) and its main feature is the strategic complementarity between the levels of effort of the different players. Coordination games typically have multiple equilibria. In this example, we can distinguish three different types of unobservables from the point of view of the outside researcher. The first type consists of payoff-relevant common knowledge unobservables (PR), e.g., classroom, school, teacher, and students characteristics that enter in the production function of students' knowledge and are known to all the players but not to the researcher. The second type consists of private information unobservables (PI), e.g., part of the students' and teacher's skills, and their respective costs of effort, are private information of these players, and they are also unknown to the researcher. Finally, in the presence

of multiple equilibria, we may have that two classes with exactly the same (payoff relevant) inputs have selected different types of equilibria. Apparently innocuous characteristics of a class may affect students' and teachers' beliefs about the effort of others. Part of these non-payoff variables affecting beliefs are unobservable to the researcher (ME unobservables). ■

Assumption 1 contains basic conditions on the structural model that are standard in the empirical literature of discrete games of incomplete information.[5]

ASSUMPTION 1. (A) Payoff functions $\{\Pi_i : i \in \mathcal{I}\}$ *are additively separable in the private information component, i.e.,* $\Pi_i = \widetilde{\pi}_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \omega) + \widetilde{\varepsilon}_i(a_i)$*, where* $\widetilde{\boldsymbol{\varepsilon}}_i \equiv \{\widetilde{\varepsilon}_i(a_i) : a_i \in \mathcal{A}\}$ *is a vector of* $J+1$ *real valued random variables. (B)* $\widetilde{\boldsymbol{\varepsilon}}_i$ *is independently distributed across players and independent of common knowledge variables* $(\mathbf{x},\omega,\xi)$ *with a distribution function that is continuously differentiable with respect to the Lebesgue measure in the Euclidean space* $\mathbb{R}^{J+1}$*.* ■

A player's optimal choice is invariant to any affine transformation of his payoff function such that we can identify the payoff function only up to an affine transformation.[6] Given a baseline choice alternative, say 0, for any $a_i \neq 0$ we define the *normalized* payoff function, $\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \omega) \equiv [\widetilde{\pi}_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \omega) - \widetilde{\pi}_i(0, \boldsymbol{a}_{-i}, \mathbf{x}, \omega)]/\delta_i$, and the *normalized* private information variables $\varepsilon_i(a_i) \equiv [\widetilde{\varepsilon}_i(a_i) - \widetilde{\varepsilon}_i(0)]/\delta_i$ where $\delta_i^2 \equiv Var(\widetilde{\varepsilon}_i(1) - \widetilde{\varepsilon}_i(0))$. For the rest of the paper we describe the model in terms of the normalized payoff functions π_i and private information variables ε_i.

ASSUMPTION 2. The researcher knows the distribution function G *of the vector of (normalized) private information variables* $\boldsymbol{\varepsilon}_i \equiv \{\varepsilon_i(a_i) : a_i \neq 0\}$*.* ■

The standard equilibrium concept in static games of incomplete information is *Bayesian Nash equilibrium* (BNE). We assume that the outcome of this game is a BNE. Under this assumption, a player's strategy is a function only of payoff-relevant variables, i.e., a function of $(\mathbf{x}, \omega, \boldsymbol{\varepsilon}_i)$. If the game has multiple equilibria, then the sunspot variables in ξ affect the selection of the equilibrium and therefore the outcome of the game. We first describe a BNE and then we incorporate the equilibrium selection mechanism when the model has multiple equilibria.

[5] In a recent working paper, Liu, Vuong, and Xu (2013) study identification of binary choice games of incomplete information relaxing the assumptions of additive separability and independence between players' private information. Wan and Xu (2014) study identification of a semiparametric binary game with correlated private information. These two papers assume that there is not common knowledge unobserved heterogeneity or multiple equilibria in the data.

[6] In this paper, we consider that the researcher has data only on players' choices and state variables. Some of our normalization assumptions can be relaxed when the researcher has data on a component of the payoff function such as firms' revenue.

Let $\boldsymbol{\sigma} = \{\sigma_i(\mathbf{x}, \omega, \boldsymbol{\varepsilon}_i) : i \in \mathcal{I}\}$ be a set of strategy functions where σ_i is a function from $\mathcal{X} \times \Omega \times \mathbb{R}^J$ into $\mathcal{A}$. Associated with a set of strategy functions we can define a vector of *conditional choice probabilities* (CCPs), $\mathbf{P}(\mathbf{x}, \omega, \boldsymbol{\sigma}) \equiv \{P_i(a_i|\mathbf{x}, \omega, \sigma_i) : (a_i, i) \in \mathcal{A} - \{0\} \times \mathcal{I}\}$ such that:

$$P_i\left(a_i \mid \mathbf{x}, \omega, \sigma_i\right) \equiv \int 1\left\{\sigma_i(\mathbf{x}, \omega, \boldsymbol{\varepsilon}_i) = a_i\right\} \, dG(\boldsymbol{\varepsilon}_i) \tag{1}$$

where $1\{.\}$ is the indicator function. These probabilities represent the expected behavior of player i from the point of view of the other players, who do not know $\boldsymbol{\varepsilon}_i$. By the independence of private information across players in Assumption 1(B), players' actions are independent once we condition on common knowledge variables $(\mathbf{x}, \omega)$ and players's strategies, such that $\Pr(a_1, a_2, ..., a_N|\mathbf{x}, \omega, \boldsymbol{\sigma}) = \prod_{i=1}^{N} P_i(a_i|\mathbf{x}, \omega, \sigma_i)$.

Given beliefs σ about the behavior of other players, each player maximizes his expected utility. Let $\pi_i^{\boldsymbol{\sigma}}(a_i, \mathbf{x}, \omega) + \varepsilon_i(a_i)$ be player i's (normalized) expected utility if he chooses alternative a_i and the other players behave according to their respective strategies in $\boldsymbol{\sigma}$. We have that:

$$\pi_i^{\boldsymbol{\sigma}}(a_i, \mathbf{x}, \omega) \equiv \sum_{\boldsymbol{a}_{-i} \in A^{N-1}} \left(\prod_{j \neq i} P_j(a_j|\mathbf{x}, \omega, \sigma_j)\right) \, \pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \omega) \tag{2}$$

DEFINITION 1: A Bayesian Nash equilibrium (BNE) in this game is a set of strategy functions $\boldsymbol{\sigma}^*$ *such that for any player* i *and for any* $(\mathbf{x}, \omega, \boldsymbol{\varepsilon}_i)$*,*

$$\sigma_i^*(\mathbf{x}, \omega, \boldsymbol{\varepsilon}_i) = \arg\max_{a_i \in \mathcal{A}} \left\{ \pi_i^{\boldsymbol{\sigma}^*}(a_i, \mathbf{x}, \omega) + \varepsilon_i(a_i) \right\} \qquad \blacksquare \tag{3}$$

We can represent a BNE in the space of choice probabilities. This representation is convenient for the econometric analysis of this model. Solving the equilibrium condition (3) into the definition of *choice probabilities* in (1) and taking into account the form of the expected payoff in (2), we can characterize a BNE as a vector of choice probabilities, $\mathbf{P}^*(\mathbf{x}, \omega) = \{P_i^*(a_i|\mathbf{x}, \omega) : a_i \neq 0, i \in I\}$, that solves the fixed point equation $\mathbf{P}^*(\mathbf{x}, \omega) = \boldsymbol{\Psi}\left(\mathbf{x}, \omega, \mathbf{P}^*(\mathbf{x}, \omega)\right)$. The fixed point mapping $\boldsymbol{\Psi}\left(\mathbf{x}, \omega, \mathbf{P}\right)$ from CCP's to CCPs is defined as $\{\Psi_i(a_i|\mathbf{x}, \omega, \mathbf{P}_{-i}) : a_i \neq 0, \, i \in I\}$, and

$$\Psi_i(a_i|\mathbf{x}, \omega, \mathbf{P}_{-i}) \equiv \int 1\left\{a_i = \arg\max_{k \in \mathcal{A}} \left(\sum_{a_{-i}} \left(\prod_{j \neq i} P_j(a_j)\right) \pi_i(k, \boldsymbol{a}_{-i}, \mathbf{x}, \omega) + \varepsilon_i(k)\right)\right\} dG(\boldsymbol{\varepsilon}_i) \tag{4}$$

We call Ψ_i *best response probability function* because it provides the probability that an action is optimal for player i given that the player believes that his opponents behave according to the probabilities in $\mathbf{P}_{-i}$.

The continuous differentiability of the distribution function $G(\boldsymbol{\varepsilon}_i)$ in Assumption 1 implies that the best response probability mapping Ψ_i is continuously differentiable in $\mathbf{P}_{-i}$. Therefore, by Brower's fixed point theorem, the mapping $\boldsymbol{\Psi}(\mathbf{x}, \omega, .)$ has at least one equilibrium. The set of equilibria associated with $(\mathbf{x}, \omega)$ is defined as $\Gamma(\mathbf{x}, \omega) \equiv \{\mathbf{P} : \mathbf{P} = \boldsymbol{\Psi}(\mathbf{x}, \omega, \mathbf{P})\}$. Under our regularity conditions, the set of equilibria $\Gamma(\mathbf{x}, \omega)$ is discrete and finite for almost all games $(\mathbf{x}, \omega, G)$. Furthermore, each equilibria belongs to a particular "type" such that a marginal perturbation in the payoff function implies a small variation in the equilibrium probabilities within the same type. The following definitions and lemma establish these results formally.

DEFINITION 2 [Singularity points and Regular BNE]. Let $f(\mathbf{x}, \omega, \mathbf{P})$ *be the function* $\mathbf{P} - \boldsymbol{\Psi}(\mathbf{x}, \omega, \mathbf{P})$ *such that an equilibrium of the game can be represented as a solution in* $\mathbf{P}$ *to the system of equations* $f(\mathbf{x}, \omega, \mathbf{P}) = 0$. *A vector* $\mathbf{P}_0$ *is a singularity point of the mapping* $f(\mathbf{x}, \omega, \mathbf{P})$ *if the Jacobian matrix* $\partial f(\mathbf{x}, \omega, \mathbf{P}_0)/\partial \mathbf{P}'$ *is singular. An equilibrium* $\mathbf{P}^*$ *is regular if and only if it is not a singularity point, i.e., if the Jacobian matrix* $\partial f(\mathbf{x}, \omega, \mathbf{P}^*)/\partial \mathbf{P}'$ *is non-singular.* ■

DEFINITION 3 [Equilibrium types]. Let $\boldsymbol{\pi}_{(\mathbf{x},\omega)} \in \mathbb{R}^{N(J+1)^N}$ *be the vector of players' payoffs associated to* $(\mathbf{x}, \omega)$. *The equilibrium mapping* $\boldsymbol{\Psi}$ *depends of* $(\mathbf{x}, \omega)$ *only through* $\boldsymbol{\pi}_{(\mathbf{x},\omega)}$ *such that we can represent the function* $f(\mathbf{x}, \omega, \mathbf{P})$ *as* $f(\boldsymbol{\pi}_{(\mathbf{x},\omega)}, \mathbf{P})$. *Let* $\boldsymbol{\pi}^0$ *and* $\boldsymbol{\pi}^1$ *be two vectors of payoffs in* $\mathbb{R}^{N(J+1)^N}$ *and let* $\mathbf{P}^{*0}$ *and* $\mathbf{P}^{*1}$ *be BNEs associated with* $\boldsymbol{\pi}^0$ *and* $\boldsymbol{\pi}^1$, *respectively. We say that* $\mathbf{P}^{*0}$ *and* $\mathbf{P}^{*1}$ *belong to the same type of equilibrium if and only if there is a continuous path* $\{\mathbf{P}[t] : t \in [0, 1]\}$ *(continuous in* t*) that satisfies the condition* $f(\ [1 - t]\ \boldsymbol{\pi}^0 + t\ \boldsymbol{\pi}^1,\ \mathbf{P}[t]) = 0$ *for every* $t \in [0, 1]$, *such that this path connects in a continuous way the equilibria* $\mathbf{P}^{*0}$ *and* $\mathbf{P}^{*1}$. ■

LEMMA 1. Under the conditions of Assumption 1: (A) For almost all payoffs $\boldsymbol{\pi}$ *the set of equilibria* $\Gamma(\boldsymbol{\pi})$ *includes only regular equilibria. (B) If the set of equilibria* $\Gamma(\boldsymbol{\pi})$ *contains only regular equilibria, then the set* $\Gamma(\boldsymbol{\pi})$ *is finite. (C) Every regular equilibrium belongs to a particular type. That is, let* $(\boldsymbol{\pi}^0, \mathbf{P}^0)$ *be a regular equilibrium. Then, there exists a neighborhood of* $\boldsymbol{\pi}^0$ *such that, for every* $\boldsymbol{\pi}$ *in that neighborhood a unique equilibrium* $\mathbf{P}[\boldsymbol{\pi}]$ *exists, and* $(\boldsymbol{\pi}$,$\mathbf{P}[\boldsymbol{\pi}])$ *is of the same type as* $(\boldsymbol{\pi}^0, \mathbf{P}^0)$. ■

Proof: In the appendix.

Based on Lemma 1, we can index equilibrium types by $\tau \in \{1, 2, ..., L_\tau\}$ and use $\Upsilon(\boldsymbol{\pi}_{(\mathbf{x},\omega)})$ to represent the set of indexes for the equilibrium types associated to a game with payoffs $\boldsymbol{\pi}_{(\mathbf{x},\omega)}$.

EXAMPLE 2: Consider a simple version of the coordination game within the classroom in Example 1. Students' choice of effort is binary: $a_i \in \{0, 1\}$. The teacher's combination of skills and effort is considered exogenous and represented by the scalar variable x. A student's payoff for choosing the high level of effort is $\pi_i(1, \boldsymbol{a}_{-i}, x) + \varepsilon_i(1)$ with $\pi_i(1, \boldsymbol{a}_{-i}, x) = \alpha + \beta\ x + \gamma\ x\ \left(\frac{1}{N-1}\sum_{j \neq i} a_j\right)$, where $\varepsilon_i(1)$ is private information (e.g., a component of the cost of effort) and it is i.i.d. across students with a standard normal distribution. All the students are assumed identical except for their private information variables. We assume that the equilibrium is symmetric , i.e., all the students have the same probability of effort $P(x)$. Then, the best response probability function of any student in this model is $\Psi(1|\ x, P) = \Phi(\alpha + \beta\ x + \gamma\ x\ P(x))$. Suppose that $x > 0$ and $\gamma > 0$ such that there are positive synergies between the teacher's effort/skills and students' effort. The model is a *coordination game* and the best response probability function has an S form as shown in Figure 1.

Figures 1 and 2 come from this example when the parameter values are $\alpha = 2.0$, $\beta = -7.31$, and $\gamma = 6.75$, and variable x is an index in the interval $[0, 1]$. Figure 1 presents the equilibrium mapping when teacher's effort is $x = 0.52$. For this value of x the model has three equilibria with low, middle, and high probability of students' effort. Figure 2 presents the equilibrium correspondence of the model. This inverted-S curve correspondence has two singularity points, at $(x, P) = (0.4757, 0.2301)$ and $(x, P) = (0.6131, 0.8447)$. At these points, the derivative $\partial[P - \Psi(x, P)]/\partial P$ is zero such that these two equilibria are not regular. For the rest of equilibria in the manifold, the derivative $\partial[P - \Psi(x, P)]/\partial P$ is different to zero such that they are all regular equilibria. The two singularity points divide the correspondence into three functions, from x into P, which correspond to the three types of equilibria of this model. For values of x in the interval $[0, 0.6131]$, the upper part of the inverted-S curve corresponds to the high-probability equilibrium type. For values of x in the interval $[0.4757, 1]$, the lower part of the curve corresponds to the low-probability equilibrium type. Finally, for values of x in the interval $[0.4757, 0.6131]$, the intermediate part of the curve corresponds to the middle equilibrium type. In this example, teacher's effort is a substitute of student's own effort in the high and low equilibrium types, i.e., the equilibrium probability of students' effort declines with teacher's effort. However, their efforts are complements in the middle equilibrium. ■

Figure 1: Coordination Game. Best Response Function $\Psi(\mathbf{x}, \mathbf{P})$
Teacher's effort: $x = 0.52$; Set of Equilibria: $\{0.054, 0.521, 0.937\}$

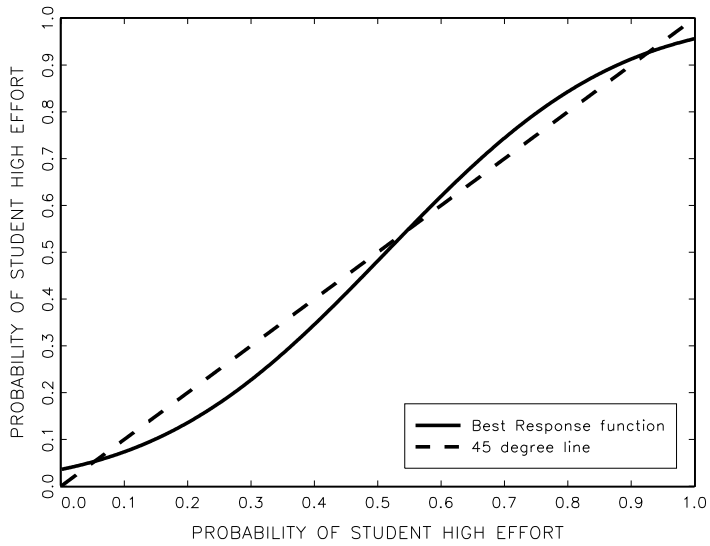
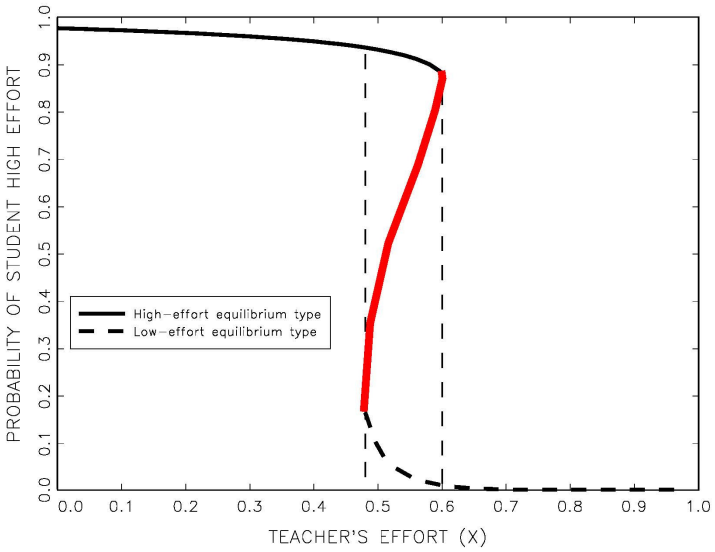


Figure 2: Coordination Game. Equilibrium Correspondence
$\{(x, P) : P = \Psi(x, P)\}$



3 Data and data generating process

Suppose that the researcher observes M different realizations of the game; e.g., M different local markets in a game of market competition. We use the index m to represent a realization of the game. For the sake of concreteness in our discussion, we consider that these multiple realizations of the game represent the same players playing the game at M different markets. For every market m, the researcher observes the vector $\mathbf{x}_m$ and players' actions $\{a_{1m},\ a_{2m},\ ...,\ a_{Nm}\}$. For the asymptotics of the estimators, we consider the case where the number of players N is small and the number of realizations of the game is large (e.g., the number of markets M goes to infinity). As stated in Assumption 2, we assume that the distribution of the normalized private information unobservables, G, is known to the researcher. We study the nonparametric identification of the normalized payoff functions π_i and of the distribution of common knowledge unobservables (ω_m, τ_m), where τ_m represents the equilibrium type selected in market m.

Let $f_\omega(\omega_m|\mathbf{x}_m)$ be the conditional probability function of ω_m given $\mathbf{x}_m$, and let $\lambda(\tau_m|\mathbf{x}_m, \omega_m)$ be the conditional probability function of τ_m given $(\mathbf{x}_m, \omega_m)$ such that $p(\tau_m, \omega_m, \mathbf{x}_m) = \lambda(\tau_m|\omega_m, \mathbf{x}_m)$ $f_\omega(\omega_m|\mathbf{x}_m)$ $p_x(\mathbf{x}_m)$. Assumption 3 summarizes all the conditions that we impose on the *Data Generating Process* (DGP).[7]

ASSUMPTION 3: (A) The realizations of the vector $(\omega_m, \tau_m, \mathbf{x}_m)$ *are independent and identically distributed across markets and independent of the private information variables* $\{\boldsymbol{\varepsilon}_{im}\}$. *(B)* $f_\omega(\omega|\mathbf{x})$ *has finite support* $\Omega \equiv \{\omega^{(1)},\ \omega^{(2)},\ ...,\ \omega^{(L_\omega(\mathbf{x}))}\}$, *i.e., finite mixture model. (C) For every value* $(\mathbf{x}, \omega)$, *all the equilibria in the DGP are regular; by Lemma 1, this implies that* $\lambda(\tau|\mathbf{x}, \omega)$ *has finite support* $\Upsilon(\boldsymbol{\pi}_{(\mathbf{x},\omega)})$. *(D) The observed vector of players actions in market* m, $\boldsymbol{a}_m \equiv \{a_{1m},\ a_{2m},\ ...,\ a_{Nm}\}$, *is a random draw from a multinomial distribution,* $\Pr(\boldsymbol{a}_m|\mathbf{x}_m, \omega_m, \tau_m) = \prod_{i=1}^N P_i^{(\tau_m)}(a_{im}|\mathbf{x}_m, \omega_m)$, *where the vector of CCPs* $\mathbf{P}^{(\tau_m)}(\mathbf{x}_m, \omega_m) \equiv \{P_i^{(\tau_m)}(a_{im}|\mathbf{x}_m, \omega_m) : (a_i, i) \in \mathcal{A} - \{0\} \times \mathcal{I}\}$ *is an equilibrium of type* τ_m, *i.e.,* $\mathbf{P}^{(\tau_m)}(\mathbf{x}_m, \omega_m) = \boldsymbol{\Psi}(\mathbf{x}_m, \omega_m, \mathbf{P}^{(\tau_m)}(\mathbf{x}_m, \omega_m))$. ■

Let $Q(\boldsymbol{a}|\mathbf{x})$ be the probability distribution of observed players' actions conditional on observed exogenous variables: $Q(\boldsymbol{a}|\mathbf{x}) \equiv \Pr(\boldsymbol{a}_m = \boldsymbol{a} \mid \mathbf{x}_m = \mathbf{x})$. This probability distribution Q is identified from the data under very mild regularity conditions. For the rest of the paper, we assume the probability function $Q(\boldsymbol{a}|\mathbf{x})$ to be known. Furthermore, this probability distribution contains all the information from the data that is relevant to identify the structural parameters of the model,

[7] Note that in the description of the DGP we do not need to specify the distribution of the vector of unobservable sunspots ξ_m but only of the selected equilibrium type τ_m.

$\{\boldsymbol{\pi},\ f_\omega,\ \lambda\}$. According to the model and our assumptions on the DGP, we have the following relationship between Q and the structural parameters $\{\boldsymbol{\pi},\ f_\omega,\ \lambda\}$:

$$\begin{aligned} Q(\boldsymbol{a}|\mathbf{x}) \quad &= \quad \sum_{\omega\in\Omega} \sum_{\tau\in\Upsilon(\boldsymbol{\pi}_{(\mathbf{x},\omega)})} f_\omega(\omega|\mathbf{x})\ \lambda(\tau|\mathbf{x},\omega)\ \left[\prod_{i=1}^{N} P_i^{(\tau)}(a_i|\mathbf{x},\omega)\right] \\ \text{subject} \quad &\text{to} \quad \mathbf{P}^{(\tau)}(\mathbf{x},\omega) = \boldsymbol{\Psi}(\mathbf{x},\omega,\mathbf{P}^{(\tau)}(\mathbf{x},\omega)) \end{aligned} \tag{5}$$

The system of equations in (5) summarizes all the restrictions imposed by the model on the data for identification of the structural parameters. Therefore, given Q, the primitive functions are identified if this system of equations has a unique solution for $\{\boldsymbol{\pi},\ f_\omega,\ \lambda\}$.

DEFINITION 4 (Identification): Suppose that the distribution Q is known to the researcher. The model is (point) identified if and only if there is a unique value $\{\boldsymbol{\pi},\ f_\omega,\ \lambda\}$ that solves the system of equations (5). ■

We are interested in two main questions: under which conditions is the payoff function identified? and under which conditions is it possible to separately identify the relative contribution of payoff-relevant (PR) and multiple-equilibria (ME) unobservables as competing explanations for non-independence of players' actions in the data?

Since the two common knowledge unobservables, ω and τ, have finite support, we can define a scalar random variable $\kappa \equiv g(\omega,\tau)$, also with finite support, that represents the combination of these two unobservables. Let $h(\kappa|\mathbf{x})$ be the PDF of κ, i.e., $h(\kappa|\mathbf{x}) = \sum_{\omega,\tau} 1\{\kappa = g(\omega,\tau)\}\ f_\omega(\omega|\mathbf{x})\ \lambda(\tau|\mathbf{x},\omega)$.

We follow a sequential approach to derive conditions for identification. In the first step, given Q, we obtain conditions for the identification of the CCPs $P_i(a_i \mid \mathbf{x},\kappa)$ and the probability distribution $h(\kappa|\mathbf{x})$ from the system of equations (i.e., nonparametric finite mixture model): $Q(\boldsymbol{a}|\mathbf{x}) = \sum_\kappa h(\kappa|\mathbf{x})\ \left[\prod_{i=1}^N P_i(a_i \mid \mathbf{x},\kappa)\right]$. Under Assumptions 1 and 2, we can apply Hotz-Miller inversion theorem (Hotz and Miller, 1993) to recover the expected payoff function of player i from the vector of CCPs of this player. Therefore, identification of the CCPs $P_i(a_i \mid \mathbf{x},\kappa)$ implies the identification of the expected payoff functions $\pi_i^{\mathbf{P}}(a_i,\mathbf{x},\kappa) \equiv \pi_i^{\boldsymbol{\sigma}}(a_i,\mathbf{x},\kappa)$ as defined in equation (2). In the second step we consider the identification of the payoff function $\pi_i(a_i,\boldsymbol{a}_{-i},\mathbf{x},\omega)$ given that the expected payoff $\pi_i^{\mathbf{P}}(a_i,\mathbf{x},\kappa)$ is known and given the system of equations (2). Finally, in step 3, we derive conditions for the identification of the distributions $f_\omega(\omega|\mathbf{x})$ and $\lambda(\tau|\mathbf{x},\omega)$ given the payoff function π_i and the distribution $h(\kappa|\mathbf{x})$.

4 Identification: Sequential approach

4.1 Model without PR or ME unobserved heterogeneity

Before we present our identification results for the model with the two sources of unobserved heterogeneity, it is helpful to discuss the identification of the model without any of these two sources of heterogeneity. This case is a useful benchmark of comparison, and it illustrates the importance of exclusion restrictions for the identification of payoffs.

Consider the model without any form of common knowledge unobserved heterogeneity, either payoff relevant or sunspots. In this restricted version of the model, ω_m is a constant across markets, and τ_m is a deterministic function of the observable $\mathbf{x}_m$, i.e., $\tau_m = f_\tau(\mathbf{x}_m)$, and the probability distribution that describes the equilibrium selection is degenerate, i.e., $\lambda(\tau|\mathbf{x}_m) = 1\{\tau = f_\tau(\mathbf{x}_m)\}$, where $1\{.\}$ is the indicator function. This condition is a *soft* version of the assumption "only one equilibrium is played in the data".

4.1.1 Step 1: Identification of equilibrium CCP's

Without common knowledge unobservables, players' actions are independent conditional on observables $\mathbf{x}$ such that $Q(\boldsymbol{a}|\mathbf{x}) = \prod_{i=1}^{N} Q_i(a_i|\mathbf{x})$ where Q_i is the marginal distribution of a_i conditional on $\mathbf{x}$. According to the model, this marginal distribution is the equilibrium CCP for player i: $P_i(a_i|\mathbf{x}, \tau = f_\tau(\mathbf{x})) = Q_i(a_i|\mathbf{x})$. If $\mathbf{x}$ has a discrete and finite support, the probabilities Q_i can be consistently estimated under very mild regularity conditions. The case of continuous variables in $\mathbf{x}$ is slightly more complicated because multiplicity of equilibria may generate discontinuity points in the CCP function. The researcher does not know, ex-ante, the number and the location of these discontinuity points, and this complicates the application of smooth nonparametric estimators, such as kernel or sieve estimators.[8] However, the discontinuity of the probability function Q does not imply that the model is not identified. Müller (1992) and Delgado and Hidalgo (2000) study nonparametric estimation of a regression function with 'change-points' or discontinuities when the location of these points is unknown to the researcher. They propose variations of standard kernel methods and show consistency and asymptotic normality.

[8] If the model has multiple equilibria, the probability function $Q_i(a_i|\mathbf{x})$ may be discontinuous in $\mathbf{x}$ if only because some equilibria can appear and disappear when we move along the space of $\mathbf{x}$. This point is illustrated in Figure 2. For any value of x in the interval $[0.48, 0.60]$, the model has multiple equilibria. However, the model has a unique equilibrium for values $x < 0.48$ or $x > 0.60$.

4.1.2 Step 2: Identification of payoffs

Given that $P_i(a_i|\mathbf{x}) = Q_i(a_i|\mathbf{x})$, we can apply Hotz-Miller inversion to uniquely recover equilibrium expected payoffs $\{\pi_i^{\mathbf{P}}(a_i, \mathbf{x}) : a_i \in \mathcal{A} - \{0\}\}$ from $\{Q_i(a_i|\mathbf{x}) : a_i \in \mathcal{A} - \{0\}\}$. We can treat expected payoffs hereafter as known. The problem of identification in step 2 is that of recovering the payoff function $\boldsymbol{\pi}$ from the system of equations:

$$\pi_i^{\mathbf{P}}(a_i, \mathbf{x}) = \sum_{\boldsymbol{a}_{-i}} P_{-i}(\boldsymbol{a}_{-i}|\mathbf{x})\ \pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}) \tag{6}$$

where $P_{-i}(\boldsymbol{a}_{-i}|\mathbf{x}) \equiv \prod_{j \neq i} P_j(a_j|\mathbf{x})$. Because of strategic interactions, there are multiple payoff values $\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x})$ for every $\pi_i^{\mathbf{P}}(a_i, \mathbf{x})$ that is identified, so a discrete game is severely under-identified relative to a standard discrete choice - random utility model. Some restrictions on payoffs are needed to restore identification.

In this literature, exclusion restrictions have been the most common type of identifying restrictions (see Bajari et al., 2010). Suppose that $\mathbf{x} = \{\mathbf{x}^c, z_i : i \in I\}$ where $z_i \in \mathcal{Z}$ and the set $\mathcal{Z}$ is discrete with at least $J + 1$ points. Furthermore, suppose that $\pi_i(a_i, a_{-i}, \mathbf{x})$ depends on $(\mathbf{x}^c, z_i)$ but not on $\mathbf{z}_{-i} \equiv \{z_j : j \neq i\}$. Then, for fixed $(\mathbf{x}^c, z_i)$ and different values of $\mathbf{z}_{-i}$ the primitive payoffs $\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x})$ on the right-hand-side of (6) are constant. However, the probabilities P_{-i} and the expected payoffs do vary with $\mathbf{z}_{-i}$ because $\mathbf{z}_{-i}$ changes the payoffs and equilibrium behavior of other players. Let $\boldsymbol{\Pi}_i^{\mathbf{P}}(a_i, \mathbf{x}^c, z_i)$ be the $|\mathcal{Z}|^{N-1} \times 1$ vector collecting $\left\{\pi_i^{\mathbf{P}}(a_i, \mathbf{x}^c, z_i, \mathbf{z}_{-i})\right\}$ for all $\mathbf{z}_{-i}$, and let $\boldsymbol{\Pi}_i(a_i, \mathbf{x}^c, z_i)$ be the $(J+1)^{N-1} \times 1$ vector collecting payoffs $\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}^c, z_i)$ for all $\boldsymbol{a}_{-i}$. Then, equations (6) can be written in vector form as

$$\boldsymbol{\Pi}_i^{\mathbf{P}}(a_i, \mathbf{x}^c, z_i) = \mathbf{P}_{-i}(\mathbf{x}^c, z_i)\ \boldsymbol{\Pi}_i(a_i, \mathbf{x}^c, z_i) \tag{7}$$

where $P_{-i}(\mathbf{x}^c, z_i)$ is a matrix with dimension $|\mathcal{Z}|^{N-1} \times (J+1)^{N-1}$ with elements $P_{-i}(\boldsymbol{a}_{-i}|\mathbf{z}_{-i}, z_i, \mathbf{x}^c)$ where each row corresponds to a different value of $\mathbf{z}_{-i}$ and each column to a different value of $\boldsymbol{a}_{-i}$. We can recover the vector of payoffs $\boldsymbol{\Pi}_i(a_i, \mathbf{x}^c, z_i)$ from (7) as long as matrix $P_{-i}(\mathbf{x}^c, z_i)$ has full column rank.

4.2 Model with both PR and ME unobserved heterogeneity

4.2.1 Step 1: Identification of equilibrium CCP's and mixing distributions

The identification of CCPs is based on the set of restrictions:

$$Q(\boldsymbol{a}|\mathbf{x}) = \sum_{\kappa=1}^{L_\kappa(\mathbf{x})} h(\kappa|\mathbf{x})\ \left[\prod_{i=1}^{N} P_i(a_i \mid \mathbf{x}, \kappa)\right] \tag{8}$$

where $L_\kappa(\mathbf{x})$ represents the number of points in the support of the distribution $h(\kappa|\mathbf{x})$. This system of equations describes a nonparametric finite mixture model. The identification of this class of models has been studied by Hall and Zhou (2003), Hall, Neeman, Pakyari and Elmore (2005), Allman, Matias, and Rhodes (2009), and Kasahara and Shimotsu (2014), among others. In all these papers, identification is based on the independence between the N variables $\{a_1, a_2, ..., a_N\}$ once we condition on $(\mathbf{x}, \kappa)$ and it does not exploit any variation in the exogenous variables in $\mathbf{x}$, e.g., independence assumptions between $\mathbf{x}$ and κ. Therefore, the analysis that follows applies separately for every value of $\mathbf{x}$ and for notational simplicity we drop $\mathbf{x}$ as an argument.

In equation (8), the necessary order condition for identification is $(J+1)^N - 1 \geqslant JNL_\kappa + (L_\kappa - 1)$, i.e., the number of restrictions or known probabilities Q should be greater or equal than the number of unknown parameters in the choice probabilities and in the distribution of the unobservables κ. The basic intuition from this order condition is that the assumption of independent marginals can deliver identification if the number of variables and/or their support are sufficiently large. Hall and Zhou (2003) studied nonparametric identification for a mixture with two branches, $L_\kappa = 2$ in our notation. They showed that the model cannot be identified for $N = 2$, even if J is made large enough to satisfy the order condition. However, for any $N \geqslant 3$ they showed that the model is generically identified (Theorem 4.3 in Hall and Zhou, 2003). Allman et al (2009) study the more general case with $L_\kappa \geq 2$ branches. They establish that a mixture with L_κ components is generically identified if $N \geqslant 3$ and $L_\kappa \leq (J+1)^{int[(N-1)/2]}$, where $int[.]$ is the *integer* or *floor* function[9]. Note that the upper bound to the number of identifiable branches not only increases with the number of variables (players) N but also with the size of support of these variables. Generic identification here means that the set of primitives for which identifiability does not hold has measure zero.

The following Proposition 1 is an application to our model of Theorem 4 and Corollary 5 in pages 13-14 of Allman et al (2009). Let $\{Y_1, Y_2, Y_3\}$ be three random variables that represent a partition of the vector of players' actions $(a_1, a_2, ..., a_N)$ such that Y_1 is equal to the action of one player (if N is odd) or two players (if N is even), and variables Y_2 and Y_3 evenly divide the actions of the rest of the players. For $j = 1, 2, 3$, let $\mathbf{P}_{Y_j}(\kappa)$ be the vector with the probability distribution of Y_j conditional on the unobserved component κ.

PROPOSITION 1. Suppose that: (a) $N \geqslant 3$*; (b)* $L_\kappa \leqslant (J+1)^{int[(N-1)/2]}$*; (c)* $h(\kappa) > 0$ *for any* $\kappa = 1, 2, ..., L_\kappa$*; and (d) for* $j = 1, 2, 3$*, the* L_κ *vectors* $\mathbf{P}_{Y_j}(\kappa = 1)$*,* $\mathbf{P}_{Y_j}(\kappa = 2)$*, ...,* $\mathbf{P}_{Y_j}(\kappa = L_\kappa)$

[9] The floor function $int[x]$ is the the greatest integer less than or equal to x.

are linearly independent. Then, the distribution h and players' CCPs P_i's are uniquely identified, up to label swapping. ■

Proof: From the proof of Theorem 4 and Corollary 5 in Allman et al (2009).

To illustrate the conditions for identification of the mixture components and weights in Proposition 1, consider the following examples. In an binary choice game with three players, the model is step 1-identified if the DGP has two mixture components, but no more. A binary choice game with five players is identified in step 1 with up to 4 mixture components, e.g., there might be a binary payoff-relevant unobservable with two different equilibria being played at each of the two values of the payoff-relevant unobservable.

In general, the true number of mixture components, L_κ, is not known by the researcher. This is particularly relevant in our model because the support of τ depends on the number of equilibria of the model that are selected in the DGP, which is an endogenous object. Therefore, it seems reasonable not to impose restrictions on the number of mixture components for κ but to identify it from the data. Kasahara and Shimotsu (2014, hereinafter KS-2014) provide conditions for identification (and estimation) of a lower bound on the number of mixture components.

Following the finite mixture literature, the number of mixture components (the *true* value of L_κ) is defined as the the smallest integer L_κ such that the finite mixture representation in equation (8) is possible.[10] First, consider a model with two players. Let $\mathbb{P}_{a_1,a_2}$ be the $(J+1)\times(J+1)$ matrix with the distribution of (a_1, a_2) such that the elements of this matrix are $\mathbb{P}_{a_1,a_2}[j,k] = \Pr(a_1 = j,\, a_2 = k)$. According to the finite mixture representation in equation (8), we have that $\mathbb{P}_{a_1,a_2} = \sum_{\kappa=1}^{L_\kappa} h(\kappa)$ $\mathbf{P}_{a_1}(\kappa)$ $\mathbf{P}_{a_2}(\kappa)'$, where $\mathbf{P}_{a_i}(\kappa)$ is the $(J+1)\times 1$ vector with the probability distribution of a_i conditional on κ. Proposition 2 in KS-2014 establishes that, generically, if $L_\kappa \leq J+1$ then $L_\kappa = rank(\mathbb{P}_{a_1,a_2})$. Therefore, for a two-players game, the rank of matrix $\mathbb{P}_{a_1,a_2}$ provides a lower bound to the number of components L_κ. For $N \geq 3$ players, Proposition 4 in KS-2014 provides a similar result. To describe this result, let a_i^* be a variable that is deterministic function of variable a_i and that *may* imply some information reduction with respect to a_i, e.g., $a_i \in \{0,1,2\}$ and $a_i^* = 1\{a_i \geq 1\}$. Let S_1 and S_2 be two random variables that come from a partition of the N variables $(a_1^*, a_2^*, ..., a_N^*)$ in two groups, e.g., $S_1 = \{a_1^*\}$ and $S_2 = \{a_2^*, ..., a_N^*\}$. Let $\widetilde{J}_1$ and $\widetilde{J}_2$ be number of points in the supports of S_1 and S_2, respectively. Let $\mathbb{P}_{S_1,S_2}$ be the $\widetilde{J}_1 \times \widetilde{J}_2$ matrix with the

[10] Given a finite mixture representation as in equation (8), it is always possible to construct other finite mixture representation with a larger value of L_κ and where the choice probabilities are linearly dependent. Therefore, we define the true number of components as the smallest value L_κ that satisfies this representation.

distribution of (S_1, S_2) such that the elements of this matrix are $\mathbb{P}_{S_1,S_2}[j,k] = \Pr(S_1 = j,\, S_2 = k)$. The finite mixture structure of the model implies that $\mathbb{P}_{S_1,S_2} = \sum_{\kappa=1}^{L_\kappa} h(\kappa)\ \mathbf{P}_{S_1}(\kappa)\ \mathbf{P}_{S_2}(\kappa)'$, where $\mathbf{P}_{S_j}(\kappa)$ is the $(J+1)\times 1$ vector with the probability distribution of S_j conditional on κ. Proposition 4 in KS-2014 establishes that, generically, if $L_\kappa \leq \min\{\widetilde{J}_1, \widetilde{J}_2\}$ then $L_\kappa = rank(\mathbb{P}_{S_1,S_2})$. The following Proposition 2 is an application to our model of Proposition 4 in Kasahara and Shimotsu (2014).

PROPOSITION 2. (A) The rank of matrix $\mathbb{P}_{S_1,S_2}$ is a lower bound of the true number of mixture components L_κ. (B) If the rank of $\mathbb{P}_{S_1,S_2}$ is strictly lower than $\min[\widetilde{J}_1, \widetilde{J}_2]$*, then the bound is tight and the number of components is exactly identified as $L_\kappa = rank(\mathbb{P}_{S_1,S_2})$.* ■

Proof: From Proposition 2 (for $N = 2$) and Proposition 4 (for $N \geq 3$) in Kasahara and Shimotsu (2014).

From Proposition 2, lower bounds on the number of mixture components are easily identifiable. Clearly, different definitions of variables S_1 and S_2 are possible and different lower bounds may be obtained depending on the researcher's choice.[11]

EXAMPLE 3: (i) Two-player game. As shown in Hall and Zhou (2003), the parameters of this model are not uniquely identified if $L_\kappa \geq 2$. However, using Proposition 2 we can identify the number of components L_κ, or at least a lower bound. With only two players we can set $S_1 = a_1$ and $S_2 = a_2$ without any data reduction, and matrix $\mathbb{P}_{S_1,S_2}$ has dimension $(J+1)\times(J+1)$. If $\mathbb{P}_{S_1,S_2}$ is full rank, then we can say that $L_\kappa \geq J+1$. Otherwise, we have that L_κ is exactly identified as the rank of $\mathbb{P}_{S_1,S_2}$. For instance, in a two-player binary choice game we have that $|\mathbb{P}_{S_1,S_2}| = Q(0,0)\ Q(1,1) - Q(1,0)\ Q(0,1)$. If this determinant is zero, then the rank of $\mathbb{P}_{S_1,S_2}$ and the value of L_κ are equal to 1. In this particular example the identification of the bound on L_κ is equivalent to the test of *no common knowledge unobserved heterogeneity* that we describe in section 3.3.4 below.

(ii) Three-player binary choice game. By Proposition 1, this model is step 1 identified if the DGP has two mixture components, but no more. Define $S_1 = \{a_1, a_2\}$ and $S_2 = \{a_3\}$ such that $\widetilde{J}_1 = 4$ and $\widetilde{J}_2 = 2$. If the rank of $\mathbb{P}_{S_1,S_2}$ is 2, then we can tell that the number of components is at least 2. If the rank of $\mathbb{P}_{S_1,S_2}$ in the data is 1 then the number of components is exactly 1 such that the model does not have unobserved heterogeneity.

[11] Section 3 in Kasahara and Shimotsu (2014) describes a fairly simple sequential algorithm for estimation of the bound based on the rank tests of Kleibergen and Paap (2006). The estimator allows the researcher to aggregate information from different choices of S_1 and S_2. Intuitively, S variables with larger supports may give more accurate lower bounds but their distributions will be estimated with less precision in any given sample than those of S variables which use data reduction.

(iii) Five-player binary choice game. This game is identified in step 1 with up to $L_\kappa = 4$ mixture components, e.g., there might be a binary payoff-relevant unobservable and two different equilibria being played at each of the two values of the payoff-relevant unobservable. In this case we can set $S_1 = (a_1, a_2, a_3)$, $S_2 = (a_4, a_5)$ and $\mathbb{P}_{S_1,S_2}$ would be 8×4. With 4 mixture components in the DGP, the rank of this matrix would be 4 and the researcher would obtain this as a lower bound on the unknown true number of components.

(iv) Five player game with three choice alternatives. The maximum number of components that can be identified is 9. If we set $a_i^* = 1(a_i \geq 1)$ for $i = 1, 2, 3$, $S_1 = (a_1^*, a_2^*, a_2^*)$ and $S_2 = (a_3, a_4)$, then $\mathbb{P}_{S_1,S_2}$ is 8×9. If the DGP had 6 components the rank of $\mathbb{P}_{S_1,S_2}$ would be 6 which is smaller than $\min[8, 9]$ so the bound is tight and the researcher would know this to be the exact number of components. ■

4.2.2 Step 2: Identification of payoff function and matching types problem

Suppose that the conditions of Propositions 1 and 2 hold such that the distribution h and the CCPs $\{P_i(a_i|\mathbf{x}, \kappa)\}$ are identified, and the number of mixture components for the unobserved heterogeneity, $L_\kappa(\mathbf{x})$, is known to the researcher.[12] Given these CCPs, we can invert the best response probability function to obtain expected payoffs $\pi_i^{\mathbf{P}}(a_i, \mathbf{x}, \kappa)$. Then, the identification of the payoff function $\boldsymbol{\pi}$ is based on the system of equations:

$$\pi_i^{\mathbf{P}}(a_i, \mathbf{x}, \kappa) = \sum_{a_{-i}} P_{-i}(\boldsymbol{a}_{-i}|\mathbf{x}, \kappa)\ \pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \omega) \tag{9}$$

where $P_{-i}(\boldsymbol{a}_{-i}|\mathbf{x}, \kappa) = \prod_{j \neq i} P_j(a_j|\mathbf{x}, \kappa)$. The researcher has not identified yet which part of the unobserved heterogeneity is PR and which part is ME. It should be clear that the worst-case scenario for the identification of the payoff function π_i is when all the unobserved heterogeneity is payoff relevant, i.e., $L_\kappa(\mathbf{x}) = L_\omega(\mathbf{x})$. Our identification strategy is agnostic but allows for this worst-case scenario. Therefore, as a working hypothesis, we allow the payoff function to depend freely on the whole unobserved component κ, i.e., $\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \kappa)$. Note that this working assumption does not introduce any bias in the estimation of the payoff function. Furthermore, once the payoff function has been recovered we will be able to identify whether for two different values of κ the payoff function is the same, and therefore these two values of κ represent variation in non-payoff-relevant unobserved heterogeneity. That procedure will be part of the identification of the probability distributions of ω and τ in step 3.

[12] Note that we allow for the number of mixtures $L_\kappa(\mathbf{x})$ to vary with the vector of exogenous observables $\mathbf{x}$.

The identification of players' payoffs is based on a similar identification argument as in the model without unobserved heterogeneity. We assume that the vector of observable state variables is $\mathbf{x} = \{\mathbf{x}^c, z_i : i \in I\}$ where, for every player i, variable z_i enters in the payoff function of this player but not in the payoffs of other players.

However, a new difficulty arises in the model with unobserved heterogeneity. As mentioned in Proposition 1, the identification of the distribution h and of CCPs is *up to label swapping*, and "pointwise" or separately for each subpopulation defined by a value of the observable $\mathbf{x}$. In order to implement the identification argument in Step 2, the researcher needs to be able to "match" mixture components which correspond to the same value of ω across different subpopulations of observables defined by the instruments. Suppose that, for every $\mathbf{x}$, every mixture component is assigned a label κ from the same set $\mathcal{K}$, e.g. a subset of natural numbers indexed by κ. A label assignment is any mapping from the set of all mixture components identified in step 1 to the set of labels $\mathcal{K}$. Suppose the researcher matches mixture components with the same label across different values of $\mathbf{x}$. If the researcher uses a label assignment which incorrectly gives the same label κ to mixture components corresponding to different values of ω, then the system of equations which exploits exclusion restrictions is not satisfied at the true payoffs, and the estimation of payoffs in step 2 will be inconsistent. Example 4 below illustrates this problem.

First we define formally a label assignment. Given identification in step 1, an unobserved type can be described in terms of the vector $\{\mathbf{x},\ h_\kappa(\mathbf{x}),\ \mathbf{P}(\mathbf{x},\kappa)\}$, e.g., the unobserved type that has probability $h_\kappa(\mathbf{x}) = 0.2$ and CCPs $\mathbf{P}(\mathbf{x},\kappa) = (0.3, 0.4, 0.1)$ when $\mathbf{x} = (5, 2, 1)$. Let $\mathcal{K}_1$ be the set with all the values of $\{\mathbf{x},\ h_\kappa(\mathbf{x}),\ \mathbf{P}(\mathbf{x},\kappa)\}$ that have been identified in step 1. A label assignment is a function from the set $\mathcal{K}_1$ into the set of labels for unobserved types.

DEFINITION 5: (A) A label assignment consists of a set of labels $\mathcal{K}_2 = \{1, 2, ..., L_\kappa^*\}$*, one label for each unobserved type, and a function* $\ell(\mathbf{x}, h_\kappa(\mathbf{x}), \mathbf{P}(\mathbf{x},\kappa))$ *from the set* $\mathcal{K}_1$ *into the set of labels* $\mathcal{K}_2$ *with the property that two different values in* $\mathcal{K}_1$ *with the same value of* $\mathbf{x}$ *should have different labels. This label assignment function determines whether unobserved types for different values of* $\mathbf{x}$ *are "matched" to the same label or not. (B) A true label assignment is one that assigns the same label to unobserved types with the same* (ω, τ) *in the DGP. (C) A label assignment is payoff-correct if it matches payoff-relevant unobservable types* ω *correctly.*

The following example illustrates the problem of *matching-unobserved-types.*

EXAMPLE 4: Consider a three-player binary choice game. Suppose that in step 1 the researcher has identified $L_\kappa = 2$ mixtures or points in the support of the unobservable κ, that we represent as κ_A and κ_B. The observable exogenous variables z_i are binary: $z_i \in \mathcal{Z} = \{0, 1\}$ for $i = 1, 2, 3$. Here we concentrate in the identification of player 1's payoff. For any value of (z_1, κ), we have a system of four equations (one for each value of (z_2, z_3)) to identify the four unknowns $\pi_1(1, \boldsymbol{a}_{-1}, z_1, \kappa)$ for $\boldsymbol{a}_{-1} \in \{(0,0), (0,1), (1,0), (1,1)\}$. For notational simplicity, in this example we omit the arguments (a_1, z_1) in the payoff functions. This system of equations is,

$$\begin{bmatrix} \pi_1^P((z_2,z_3)=(0,0),\kappa) \\ \pi_1^P((z_2,z_3)=(0,1),\kappa) \\ \pi_1^P((z_2,z_3)=(1,0),\kappa) \\ \pi_1^P((z_2,z_3)=(1,1),\kappa) \end{bmatrix} =$$

$$\begin{bmatrix} P_{-1}(0,0|0,0,\kappa) & P_{-1}(0,1|0,0,\kappa) & P_{-1}(1,0|0,0,\kappa) & P_{-1}(1,1|0,0,\kappa) \\ P_{-1}(0,0|0,1,\kappa) & P_{-1}(0,1|0,1,\kappa) & P_{-1}(1,0|0,1,\kappa) & P_{-1}(1,1|0,1,\kappa) \\ P_{-1}(0,0|1,0,\kappa) & P_{-1}(0,1|1,0,\kappa) & P_{-1}(1,0|1,0,\kappa) & P_{-1}(1,1|1,0,\kappa) \\ P_{-1}(0,0|1,1,\kappa) & P_{-1}(0,1|1,1,\kappa) & P_{-1}(1,0|1,1,\kappa) & P_{-1}(1,1|1,1,\kappa) \end{bmatrix} \begin{bmatrix} \pi_1(a_{-1}=(0,0),\kappa) \\ \pi_1(a_{-1}=(0,1),\kappa) \\ \pi_1(a_{-1}=(1,0),\kappa) \\ \pi_1(a_{-1}=(1,1),\kappa) \end{bmatrix} \tag{10}$$

Given an assignment of unobserved types across the different values of $\mathbf{z}_{-1}$, the researcher constructs the vector $\pi_1^{\mathbf{P}}(\kappa)$ and the matrix $P_{-1}(\kappa)$ and solves in equation (10) for the vector of payoffs as $\pi_1(\kappa) = [\mathbf{P}_{-1}(\kappa)]^{-1} \pi_1^{\mathbf{P}}(\kappa)$. A key condition for the consistency of this estimator is that the matching of unobserved types is correct. Table 1 presents a numerical example. Panel I illustrates the case when the researcher makes a correct matching of unobserved types such that the estimator of payoffs is consistent. Panel II presents the case when the researcher makes a correct assignment of unobserved types for $(z_2, z_3) = (0, 0)$ and $(z_2, z_3) = (0, 1)$, but for values $(z_2, z_3) = (1, 0)$ and $(z_2, z_3) = (1, 1)$ the researcher swaps the correct types. Therefore, in the estimation of payoffs the researcher solves the incorrect system of equations. In the system of equations for κ_A the two bottom rows come incorrectly from $\pi_1^{\mathbf{P}}(\kappa_B)$ and $P_{-1}(\kappa_B)$, and the opposite occurs in the system of equations for κ_B. We see that the estimated payoffs are very seriously biased for the two unobserved types. The bias is not just in the level or/and the scale of the payoffs but the whole pattern of strategic interactions is inconsistently estimated. ■

Table 1
Matching unobservable types across different values of the instruments

Panel I: Consistent Matching

	Unobserved type κ_A					
	Probs $P_{-1}(a_{-1} \mid z_{-1}, \kappa_A)$				Estimated $\pi_1(a_{-1}, \kappa_A)$	
z_{-1}	$P_2(1\|z_{-1}, \kappa_A)$	$P_3(1\|z_{-1}, \kappa_A)$	$\pi_1^P(z_{-1}, \kappa_A)$	a_{-1}	$[P_{-1}(\kappa_A)]^{-1}\pi_1^P(\kappa_A)$	True $\pi_1(a_{-1}, \kappa_A)$
(0,0)	0.70	0.60	1.04	(0,0)	6.0	6.0
(0,1)	0.50	0.75	1.00	(0,1)	2.0	2.0
(1,0)	0.90	0.45	0.96	(1,0)	2.0	2.0
(1,1)	0.80	0.70	0.42	(1,1)	-1.0	-1.0
	Unobserved type κ_B					
	Probs $P_{-1}(a_{-1} \mid z_{-1}, \kappa_B)$				Estimated $\pi_1(a_{-1}, \kappa_B)$	
z_{-1}	$P_2(1\|z_{-1}, \kappa_B)$	$P_3(1\|z_{-1}, \kappa_B)$	$\pi_1^P(z_{-1}, \kappa_B)$	a_{-1}	$[P_{-1}(\kappa_B)]^{-1}\pi_1^P(\kappa_B)$	True $\pi_1(a_{-1}, \kappa_B)$
(0,0)	0.20	0.15	1.03	(0,0)	2.0	2.0
(0,1)	0.10	0.50	-0.15	(0,1)	0.0	0.0
(1,0)	0.60	0.05	0.63	(1,0)	-2.0	-2.0
(1,1)	0.50	0.40	-0.40	(1,1)	-3.0	-3.0

Panel II: Inconsistent Matching

	Unobserved type κ_A [* represents incorrect matching]					
	Probs $P_{-1}(a_{-1} \mid z_{-1}, \kappa_A)$				Estimated $\pi_1(a_{-1}, \kappa_A)$	
z_{-1}	$P_2(1\|z_{-1}, \kappa_A)$	$P_3(1\|z_{-1}, \kappa_A)$	$\pi_1^P(z_{-1}, \kappa_A)$	a_{-1}	$[P_{-1}(\kappa_A)]^{-1}\pi_1^P(\kappa_A)$	True $\pi_1(a_{-1}, \kappa_A)$
(0,0)	0.70	0.60	1.04	(0,0)	-15.1	6.0
(0,1)	0.50	0.75	1.00	(0,1)	11.1	2.0
(1,0)	0.60*	0.05*	0.63*	(1,0)	8.1	2.0
(1,1)	0.50*	0.40*	-0.40*	(1,1)	-4.1	-1.0
	Unobserved type κ_B [* represents incorrect matching]					
	Probs $P_{-1}(a_{-1} \mid z_{-1}, \kappa_B)$				Estimated $\pi_1(a_{-1}, \kappa_B)$	
z_{-1}	$P_2(1\|z_{-1}, \kappa_B)$	$P_3(1\|z_{-1}, \kappa_B)$	$\pi_1^P(z_{-1}, \kappa_B)$	a_{-1}	$[P_{-1}(\kappa_B)]^{-1}\pi_1^P(\kappa_B)$	True $\pi_1(a_{-1}, \kappa_B)$
(0,0)	0.20	0.15	1.03	(0,0)	1.4	2.0
(0,1)	0.10	0.50	-0.15	(0,1)	1.6	0.0
(1,0)	0.90*	0.45*	0.96*	(1,0)	-2.0	-2.0
(1,1)	0.80*	0.70*	0.42*	(1,1)	0.4	-3.0

The problem of *matching-unobserved-types* in the identification of payoffs in step 2 may appear not only in games but also in single-agent models, static or dynamic, with a single equilibrium. Example 5 illustrates this problem in the context of a single-agent decision model.

EXAMPLE 5: Consider a single-agent decision model where the payoff function of agent i is $\pi(a_i, \mathbf{x}_i, \omega_i) + \varepsilon_i(a_i)$. The researcher has panel data on $\{a_{it}, \mathbf{x}_{it}\}$ from many agents over a short period of time. The unobservable ω_i is time invariant and has a nonparametric finite mixture

distribution. The unobservables ε_{it} are i.i.d. over time and independent of $(\mathbf{x}_{it}, \omega_i)$. In step 1, with $T \geq 3$, the CCP function $P(a_i|\mathbf{x}, \omega)$ is nonparametrically identified, up to label swapping. Applying Hotz-Miller inversion, we can identify nonparametrically, up to label swapping, the payoff function $\pi(a, \mathbf{x},\omega)$. The identification of $\pi(a, \mathbf{x},\omega)$ up to label swapping is not a problem if the researcher is interested in the variation of payoffs with respect to the choice variable a keeping both $\mathbf{x}$ and ω constant, or if she is interested in the average payoff given $(a, \mathbf{x})$ and integrated over the distribution of ω. However, label swapping creates an identification problem if the researcher is interested in the identification of the ceteris paribus effect of a change in $\mathbf{x}$ keeping ω constant (or viceversa). The researcher does not know how to keep the unobserved type constant when the observable variables $\mathbf{x}$ vary.

Label swapping creates also an identification problem if the researcher wants to impose semiparametric restrictions on the payoff function. Suppose that the vector of explanatory variables has two components, $\mathbf{x} = (\mathbf{z}, \mathbf{w})$, and the payoff function has the following semiparametric structure, $\pi(a, \mathbf{x},\omega) = f(a, \mathbf{z})'\theta + g(a, \mathbf{w},\omega)$ where $f(a, \mathbf{z})$ is a vector of known functions, θ is a vector of unknown parameters, and $g(a, \mathbf{w},\omega)$ is a nonparametric function. To identify θ we need to fix the value of $g(a, \mathbf{w},\omega)$ and construct a system of equations for different values of $\mathbf{z}$. This identification is subject to the matching-types problem because the researcher should be able to keep the unobserved type ω constant for different values of the observable $\mathbf{z}$. ■

Therefore, without further assumptions, step-2 identification requires a label assignment that matches payoff-relevant unobservable types ω correctly across different values of the observable variables $\mathbf{x}$. We call this a *payoff-correct* assignment, or a *correct* assignment in short. The *true label assignment* assigns the same label to mixture components with the same (ω, τ) in the DGP. The true label assignment is (of course) payoff-correct. An assignment which is not the true one is still payoff-correct if it matches correctly the mixture components corresponding to the same value of the payoff-relevant unobservable ω, but not the mixture components corresponding to different equilibrium types. Note that if multiple equilibria is the only source of unobserved heterogeneity, then all assignments are payoff-correct. On the contrary, if all unobserved heterogeneity is payoff-relevant only the true assignment is correct. Because the problem of matching-unobserved-types, identification in step 2 requires an extended rank condition. Consider the vector form representation of the system of equations (9):

$$\mathbf{\Pi}_i^{\mathbf{P}}(a_i, \mathbf{x}^c, z_i, \kappa) = \mathbf{P}_{-i}(\mathbf{x}^c, z_i, \kappa)\ \mathbf{\Pi}_i(a_i, \mathbf{x}^c, z_i, \kappa) \tag{11}$$

where the vectors $\mathbf{\Pi}_i^{\mathbf{P}}$ and $\mathbf{\Pi}_i$ and the matrices P_{-i} have the same dimension and interpretation as in equation (7) but now they are conditional on the unobservable κ. For any correct assignment, this system of equations holds at the true vector of payoffs $\mathbf{\Pi}_i$. In general, for any incorrect assignment, the system does not hold at the true vector of payoffs. This is regardless the matrix P_{-i} is full rank or not. We say that a label assignment is *consistent* if it is payoff-correct and the matrix P_{-i} is full column rank for any player i. By definition, for a *consistent label assignment* the system has a unique solution and this solution provides the true payoffs. Clearly, the set of consistent assignments is included in the set of correct assignments. The set of correct assignments is non-empty because it always includes the true assignment. Instead, the set of consistent assignments can be empty if for every correct assignment the matrix P_{-i} is not full column rank.

Two sets of necessary and sufficient conditions should be satisfied in order to achieve step 2 identification. First, the set of consistent assignments should be non-empty. This is a necessary condition for identification but it is not sufficient. Suppose that there is an incorrect assignment for which the system of equations has a solution. This solution is different to the true payoffs but the researcher cannot distinguish between the solution from a consistent assignment and the solution from an incorrect assignment. To avoid this under-identification, we need to impose the condition that the system does not have a solution for any of the incorrect assignments. Proposition 3 establishes formally these necessary and sufficient conditions for identification in step 2.

PROPOSITION 3: Under exclusion restrictions, the model is identified in step 2 if and only if: (a) the set of consistent label assignments is non-empty, i.e., there exists one payoff-correct label assignment for which matrix $P_{-i}(\mathbf{x}^c, z_i, \kappa)$ has full column rank for every player i; and (b) for every payoff-incorrect label assignment, there is at least one player i for which the rank of the augmented matrix $[\mathbf{\Pi}_i^{\mathbf{P}}(a_i, \mathbf{x}^c, z_i, \kappa) \mid P_{-i}(\mathbf{x}^c, z_i, \kappa)]$ is larger than the rank of matrix $P_{-i}(\mathbf{x}^c, z_i, \kappa)$, i.e., the system of equations does not have a solution. ■

Proof: By contradiction and the application of Rouché-Capelli theorem. As explained above, conditions (a) and (b) are sufficient to obtain step 2 identification. If (a) does not hold, it is clear that we cannot recover the true vector of payoffs. If (b) does not hold, there is an incorrect assignment and a vector of payoffs different to the true payoffs that solve the system of equations. The researcher has two different solutions and cannot distinguish which one is the true. ■

Condition (a) is like the standard rank condition of the model with no heterogeneity. Condition (b) rules out that incorrect label assignments and payoffs can explain the data. Unlike the rank condition for the model with no heterogeneity, the extended rank conditions of Proposition 3 are not testable. Because the number of label assignments is finite, the researcher can sweep through all of them, check the associated system of step 2 restrictions, and more specifically, construct the set S^* of assignments for which the system has a unique solution, i.e., for every player i, P_{-i} is full column rank. However, conditions (a) and (b) are not jointly testable because the researcher does not know ex-ante whether a particular label assignment is correct or incorrect. For instance, suppose that the researcher sweeps all the possible label assignments and finds that there is only one assignment in set S^*. Without further restrictions, the researcher does not know if this assignment is payoff-correct, such that conditions (a) and (b) hold and the model is step 2 identified, or if the assignment is incorrect, such that neither condition (a) nor condition (b) hold and there is not step 2 identification.

However, based on sweeping through all the possible assignments and the construction of set S^*, it is possible to obtain a testable necessary condition for step 2 identification. Proposition 4 establishes this identification result.

PROPOSITION 4: Suppose that exclusion restrictions hold. Define S^* as the set of label assignments for which the system (11) has a unique solution, i.e., for every player i, P_{-i} is full column rank. A necessary condition for step 2 identification is that S^* is non-empty and all the assignments within this set imply the same solution. This identification condition is testable. ■

Proof: From Proposition 3, there is step 2 identification if and only if conditions (a) and (b) hold. It is clear that if conditions (a) and (b) hold, then set S^* is not empty and all the assignments within this set imply the same solution. ■

The identification conditions in Propositions 3 and 4 are not on the primitives of the model. In some applications, the researcher may be interested in imposing restrictions on the primitives of the model that imply step 2 identification. Here we present two different sufficient conditions on the primitives that provide identification of payoffs in step 2: (i) ranking independence between unobservables and observables; and (ii) additive separability and mean independence of the unobservables.

(i) Ranking independence. Suppose that the distribution of κ may vary with $\mathbf{x}$ but this dependence

does not affect the ranking of unobserved types according to the values of the probabilities $h(\kappa|\mathbf{x})$. That is, for any value of $\mathbf{x}$ we have that $h(\kappa^{(1)}|\mathbf{x}) > h(\kappa^{(2)}|\mathbf{x}) > ... > h(\kappa^{(L_\kappa)}|\mathbf{x})$. Under this condition, we can use the ranking of probabilities to match correctly unobserved types. For any value $\mathbf{x}$, type $\kappa^{(1)}$ corresponds to the highest probability, type $\kappa^{(2)}$ to the second largest probability, and so on. This restriction is weaker than independence, it still restricts the number of unobserved types L_κ to be constant across different values of $\mathbf{x}$. This particular restriction is testable after the identification of $h(\kappa|\mathbf{x})$ in step 1.

(ii) Additive separability. Suppose that the payoff function $\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \omega)$ is additively separable between the unobservable ω and the opponents' actions $\boldsymbol{a}_{-i}$:

$$\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \omega) = \beta_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}) + \eta_i(a_i, \mathbf{x}, \omega) \tag{12}$$

Furthermore, for any value of $(a_i, \mathbf{x})$ the random variable $\eta_i(a_i, \mathbf{x}, \omega)$ is mean independent of $\mathbf{x}$ and, without loss of generality, this mean is zero, i.e., $E_{\omega|\mathbf{x}}(\eta_i(a_i, \mathbf{x}, \omega)) = 0$. We also assume that the component $\beta_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x})$ of the payoff function satisfies the standard exclusion restriction such that (with some abuse of notation) $\beta_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}) = \beta_i(a_i, \boldsymbol{a}_{-i}, z_i, \mathbf{x}^c)$. Under these conditions, equation (9) has the following form:

$$\pi_i^{\mathbf{P}}(a_i, \mathbf{x}, \kappa) = \eta_i(a_i, \mathbf{x}, \omega) + \sum_{\boldsymbol{a}_{-i}} P_{-i}(\boldsymbol{a}_{-i}|\mathbf{x}, \kappa)\ \beta_i(a_i, \boldsymbol{a}_{-i}, z_i, \mathbf{x}^c) \tag{13}$$

From the identification in step 1 the researcher knows the probability distribution $h(\kappa|\mathbf{x})$ up to label swapping. Using this information we can construct a version of equation (13) integrated over the distribution of $h(\kappa|\mathbf{x})$. Note that this integration is not subject to any problem of matching-types, i.e., for any fixed $\mathbf{x}$, the distribution $h(\kappa|\mathbf{x})$ is known. This integrated equation is:

$$\overline{\pi}_i^{\mathbf{P}}(a_i, \mathbf{x}) = \sum_{\boldsymbol{a}_{-i}} \overline{P}_{-i}(\boldsymbol{a}_{-i}|\mathbf{x})\ \beta_i(a_i, \boldsymbol{a}_{-i}, z_i, \mathbf{x}^c) \tag{14}$$

where $\overline{\pi}_i^{\mathbf{P}}(a_i, \mathbf{x}) \equiv \sum_\kappa h(\kappa|\mathbf{x})\ \pi_i^{\mathbf{P}}(a_i, \mathbf{x}, \kappa)$, $\overline{P}_{-i}(\boldsymbol{a}_{-i}|\mathbf{x}) \equiv \sum_\omega h(\kappa|\mathbf{x})\ P_{-i}(\boldsymbol{a}_{-i}|\mathbf{x}, \omega)$, and we have used the condition $E_{\kappa|\mathbf{x}}(\eta_i(a_i, \mathbf{x}, \omega)) = E_{\omega|\mathbf{x}}(\eta_i(a_i, \mathbf{x}, \omega)) = 0$. Note that functions $\overline{\pi}_i^{\mathbf{P}}(a_i, \mathbf{x})$ and $\overline{P}_{-i}(\boldsymbol{a}_{-i}|\mathbf{x})$ are identified from step 1 and, very importantly, they are not subject to the matching-types problem. Given equation (14), the payoff function $\beta_i(.)$ is identified under similar conditions as the model without unobserved heterogeneity. We can construct a matrix $\overline{\mathbf{P}}_{-i}(z_i, \mathbf{x}^c)$ with elements $\overline{P}_{-i}(\boldsymbol{a}_{-i}|\mathbf{z}_{-i}, z_i, \mathbf{x}^c)$ where each row corresponds to a different value of $\mathbf{z}_{-i}$ and each column to a different value of $\boldsymbol{a}_{-i}$. Function β_i is identified if matrix $\overline{\mathbf{P}}_{-i}(z_i, \mathbf{x}^c)$ is full column rank for any value of $(z_i, \mathbf{x}^c)$.

The restriction of additive separability can be relaxed if the interaction between ω and $\boldsymbol{a}_{-i}$ in the payoff function has a parametric form. For instance, suppose that the payoff function has the following form, $\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \omega) = \beta_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x})\ g(\omega) + \eta_i(a_i, \mathbf{x}, \omega)$, where $g(.)$ is a parametric function that is perfectly known to the researcher, e.g., $g(\omega) = \exp\{\omega\}$. It is straightforward to extend the previous identification argument to this semiparametric model.[13]

4.2.3 Step 3: Identification of distributions for the two types of heterogeneity

Suppose that the conditions in Propositions 1 and 3 hold such that the researcher has identified the distribution $h(\kappa|\mathbf{x})$ and the payoff functions π_i. Now, we want to identify the probability distributions $f_\omega(\omega|\mathbf{x})$ and $\lambda(\tau|\mathbf{x}, \omega)$. There are two sets of restrictions that we can exploit to identify these distributions: (1) the payoff π_i depends on ω but not on τ; and (2) by definition, $h(\kappa|\mathbf{x}) = 1\{\kappa = g(\omega, \tau)\}\ f_\omega(\omega|\mathbf{x})\ \lambda(\tau|\mathbf{x}, \omega)$.

Let $\mathbf{\Pi}_i(\mathbf{x})$ be the matrix with dimension $J(J+1)^{N-1} \times L_\kappa(\mathbf{x})$ that contains all the payoffs $\{\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \kappa)\}$ for a given value of $\mathbf{x}$. More specifically, each column corresponds to a value of κ and it contains the payoffs $\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \kappa)$ for every value of $(a_i, \boldsymbol{a}_{-i})$ with $a_i > 0$. If two values of κ represent the same value of ω, then the corresponding columns in the matrix $\mathbf{\Pi}_i(\mathbf{x})$ should be equal. Therefore, the number of distinct columns in the payoff matrix $\mathbf{\Pi}_i(\mathbf{x})$ should be equal to $L_\omega(\mathbf{x})$. That is, we can identify the number of mixtures $L_\omega(\mathbf{x})$ as:

$$L_\omega(\mathbf{x}) = \text{Number of distinct columns in } \mathbf{\Pi}_i(\mathbf{x}) \tag{15}$$

The information in matrix $\mathbf{\Pi}_i(\mathbf{x})$ not only identifies the number of points in the support of the PR unobservables ω, but it also identifies the inverse of the mapping $\kappa = g(\omega, \tau)$ such that we know the value of (ω, τ) that corresponds to each value of κ. We use $\omega(\kappa)$ and $\tau(\kappa)$ to represent this inverse mapping. Without loss of generality we can make $\tau(\kappa) = \kappa$ for every κ.[14] We sweep through the different columns of $\mathbf{\Pi}_i(\mathbf{x})$ (i.e., the different of κ): (a) if two columns, say κ and κ', are equal, then we assign them the same value ω, i.e., $\omega(\kappa) = \omega(\kappa')$; (b) if the columns are different, then $\omega(\kappa) \neq \omega(\kappa')$.

[13] With a nonparametric payoff function that does not impose additive separability, the equation integrated over the distribution of ω is similar to (14) but includes an additional term that depends on the covariance between $P_{-i}(\boldsymbol{a}_{-i}|\mathbf{x}, \kappa)$ and $\beta_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \kappa)$ conditional on $(a_i, \boldsymbol{a}_{-i}, \mathbf{x})$. This covariance term is not zero and it is unknown to the researcher. This implies that the average payoff is not identified from this integrated equation.

[14] Note that at this stage, without solving the equilibrium (and applying an homotopy method) we cannot establish whether two vectors of CCPs for two different values of ω correspond to the same equilibrium type or not. Therefore, for the moment we consider that they are different equilibrium types.

Given the identification of the mappings $\omega(\kappa)$ and $\tau(\kappa)$, the probability distribution of the payoff relevant heterogeneity, $f_\omega(\omega|\mathbf{x})$, is identified as:

$$f_\omega(\omega|\mathbf{x}) = \sum_{\kappa=1}^{L_\kappa(\mathbf{x})} 1\{\omega(\kappa) = \omega\} \; h(\kappa|\mathbf{x}) \tag{16}$$

Taking into account that $\tau(\kappa) = \kappa$, the probability distribution of multiple equilibria heterogeneity, $\lambda(\tau|\mathbf{x}, \omega)$, is identified as:

$$\lambda(\tau|\mathbf{x}, \omega) = \frac{h(\tau|\mathbf{x})}{\sum_{\kappa=1}^{L_\kappa(\mathbf{x})} 1\{\omega(\kappa) = \omega(\tau)\} \; h(\kappa|\mathbf{x})} \tag{17}$$

PROPOSITION 5. Under the conditions of Propositions 1 and 3, the probability distributions of the unobservables, $f_\omega(\omega|\mathbf{x})$ and $\lambda(\tau|\mathbf{x}, \omega)$, are nonparametrically identified. ■

EXAMPLE 6. Suppose that $L_\kappa(\mathbf{x}) = 7$ such that $\mathbf{\Pi}_i(\mathbf{x})$ has seven columns that we label as $\kappa = 1, 2, ..., 7$. Suppose the number of distinct columns of $\mathbf{\Pi}_i(\mathbf{x})$ is 4 such that $L_\omega(\mathbf{x}) = 4$. Columns 1, 2, and 4 are equal to each other, and columns 6 and 7 are also equal to each other. Then, we have that $f_\omega(1|\mathbf{x}) = h(1|\mathbf{x}) + h(2|\mathbf{x}) + h(4|\mathbf{x})$, with $\lambda(\tau|\mathbf{x}, \omega = 1) = \frac{h(\tau|\mathbf{x})}{h(1|\mathbf{x})+h(2|\mathbf{x})+h(4|\mathbf{x})}$ for $\tau \in \{1, 2, 4\}$; $f_\omega(2|\mathbf{x}) = h(3|\mathbf{x})$, with $\lambda(3|\mathbf{x}, \omega = 2) = 1$; $f_\omega(3|\mathbf{x}) = h(5|\mathbf{x})$, with $\lambda(5|\mathbf{x}, \omega = 3) = 1$; and $f_\omega(4|\mathbf{x}) = h(6|\mathbf{x}) + h(7|\mathbf{x})$, with $\lambda(\tau|\mathbf{x}, \omega = 4) = \frac{h(\tau|\mathbf{x})}{h(6|\mathbf{x})+h(7|\mathbf{x})}$ for $\tau \in \{6, 7\}$. ■

4.3 Testable restrictions on unobserved heterogeneity

(i) Testing null hypothesis of no common knowledge unobserved heterogeneity. The model without PR and ME unobservables imposes the restriction that players' actions are independent conditional on the observable $\mathbf{x}$: $Q(\boldsymbol{a}|\mathbf{x}) = \prod_{i=1}^{N} Q_i(a_i|\mathbf{x})$. This assumption can be easily tested using a test of the null hypothesis of independence. For instance, for a binary choice game with two players the testable restriction is:

$$Q(1, 1|\mathbf{x}) \; Q(0, 0|\mathbf{x}) = Q(1, 0|\mathbf{x}) \; Q(0, 1|\mathbf{x}) \tag{18}$$

(ii) Testing null hypothesis of no ME unobserved heterogeneity. If there is not ME unobserved heterogeneity, then the number of points in the support of ω, $L_\omega(\mathbf{x})$, should be equal to the points of support of κ for any value of $\mathbf{x}$ in the sample. Therefore, taking into account that $L_\kappa(\mathbf{x}) = cols(\mathbf{\Pi}_i(\mathbf{x}))$ and that $L_\omega(\mathbf{x}) = distinct\_cols(\mathbf{\Pi}_i(\mathbf{x}))$, testing for the null hypothesis of *"no ME unobserved heterogeneity"* is equivalent to testing for:

$$\text{For every value of } \mathbf{x}, \; cols(\mathbf{\Pi}_i(\mathbf{x})) = distinct\_cols(\mathbf{\Pi}_i(\mathbf{x})). \tag{19}$$

(iii) Testing null hypothesis of no PR unobserved heterogeneity. If there is not PR unobserved heterogeneity, then for any value of $\mathbf{x}$ in the sample the number of points in the support of ω should be equal to the 1. This implies that testing for the null hypothesis of *"no PR unobserved heterogeneity"* is equivalent to testing for:

$$\text{For every value of } \mathbf{x},\ distinct\_cols(\mathbf{\Pi}_i(\mathbf{x})) = 1. \tag{20}$$

Therefore, the tests for these null hypotheses can be described in terms of tests of the rank of a matrix of statistics. They can be implemented using, for instance, the rank tests proposed by Kleibergen and Paap (2006).

Our identification results and tests rely importantly on our model specification and assumptions, e.g. independence of private information unobservables, a particular information structure of the game and the equilibrium concept of BNE. In our model, unobservables are either common-knowledge to all the players or privately known by only one player. Some departures from our specification of the information structure can invalidate our identification results. For instance, if private information unobservables are not player-specific but shared by a subgroup of players, our current step-1 identification, and for that matter the identification of the whole model, would not be valid.

5 Joint identification

All the previous identification results are based on the sequential approach. The exclusion restrictions we exploit in step 2 are quite natural in the estimation of games, and they are necessary for nonparametric identification even in games without common knowledge unobserved heterogeneity. However, the conditions for the identification of the nonparametric finite mixture in step 1 are more stringent and rule out some interesting applications, e.g., two-player games. An important question is whether these restrictions are really necessary for identification. More precisely, suppose that we do not follow a sequential approach to identify/estimate the model but we estimate jointly all the structural functions: is it possible to obtain identification even when the conditions in Proposition 1 are not satisfied? In this section we study this issue. We do this by comparing rank conditions for sequential and joint identification. We provide some insights into the relationship between sequential and joint identification. and we show that when the exclusion restrictions that are needed to identify the payoff function in step 2 provide over-identifying restrictions, these can help identify

the mixture components even when Step 1 identification conditions are not satisfied.

It is important to underline that, while in section 4 we present conditions for global (sequential) identification, our analysis below deals with local identification, both sequential and joint. We are particularly interested in showing that there is a class of models where local step-1 identification fails, yet local joint identification holds. We characterize this class in terms of rank conditions. In section 5.4, we use an example to illustrate that this class is not empty. Because local identification is necessary for global identification, for all the models within this class the sufficient conditions for step 1 identification of Proposition 1 fail.[15]

As we have shown above, once the mixing distributions h and the payoff vectors $\boldsymbol{\pi}$ have been identified, disentangling PR and ME heterogeneity in step 3 does not require any additional assumptions. Therefore, our discussion of sequential versus joint identification concentrates on steps 1 and 2.

Define the vectors of parameters: $\mathbf{h}$, with the mixing probabilities $h(\kappa|\mathbf{x})$ for every value of $(\kappa, \mathbf{x})$; $\mathbf{P}$, with the choice probabilities $P_i(a_i \mid \mathbf{x}, \kappa)$ for every player and value of $(a_i, \mathbf{x}, \kappa)$; and $\boldsymbol{\pi}$, with the payoffs $\pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \kappa)$ for every player and value of $(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \kappa)$. Let $(\mathbf{h}^0, \mathbf{P}^0, \boldsymbol{\pi}^0)$ be the true value of $(\mathbf{h}, \mathbf{P}, \boldsymbol{\pi})$ in the population. We are interested in the point identification of $(\mathbf{h}^0, \mathbf{P}^0, \boldsymbol{\pi}^0)$.[16] To compare the rank conditions for identification under the sequential and joint approaches, it is convenient to describe the identification problem as a constrained maximum likelihood problem in the population. The model implies that $\Pr(\boldsymbol{a}|\ \mathbf{x}, \mathbf{h}, \mathbf{P}) = \sum_{\kappa=1}^{L_\kappa(\mathbf{x})} h(\kappa|\mathbf{x})\ \prod_{i=1}^{N} P_i(a_i \mid \mathbf{x}, \kappa)$. The (population) log-likelihood function for the actions $\boldsymbol{a}$ conditional on $\mathbf{x}$ and parameters $(\mathbf{h}, \mathbf{P})$ is:

$$\ell(\mathbf{h}, \mathbf{P}) \quad = \quad \sum_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}) \sum_{\boldsymbol{a} \in \mathcal{A}^N} Q(\boldsymbol{a}|\mathbf{x})\ \ln \left[\sum_{\kappa=1}^{L_\kappa(\mathbf{x})} h(\kappa|\mathbf{x})\ \left(\prod_{i=1}^{N} P_i(a_i \mid \mathbf{x}, \kappa) \right) \right] \tag{21}$$

We use $\nabla \ell^0$ and $\nabla^2 \ell^0$ to represent the gradient vector and the Hessian matrix, respectively, of the likelihood function evaluated at the true value $(\mathbf{h}^0, \mathbf{P}^0)$.

5.1 Rank condition for sequential identification

5.1.1 Rank condition for step 1 identification

By the information inequality, the true value $(\mathbf{h}^0, \mathbf{P}^0)$ maximizes the likelihood $\ell(\mathbf{h}, \mathbf{P})$. We say that $(\mathbf{h}^0, \mathbf{P}^0)$ is point identified in step 1 (up to label swapping) if it uniquely maximizes this likelihood.

[15] It is also important to keep in mind that: (i) the conditions in Proposition 1 for global step-1 identification are sufficient; (ii) the conditions in Proposition 3 for global step-2 identification are necessary and sufficient; and (iii) the conditions that we present below for local identification (either step-1, or step-2, or joint) are necessary and sufficient.

[16] We assume that $L_\kappa(\mathbf{x})$ is known for all $\mathbf{x}$, e.g. by use of Proposition 2.

It is clear that $\ell(\mathbf{h}, \mathbf{P})$ is twice continuously differentiable with respect to $(\mathbf{h}, \mathbf{P})$. Suppose that $(\mathbf{h}^0, \mathbf{P}^0)$ is an interior point in the probability space such that every probability in these vectors is strictly greater than zero. Then, $(\mathbf{h}^0, \mathbf{P}^0)$ should satisfy the first order conditions of optimality $\nabla \ell^0 = 0$. The rank condition for local identification in step 1 is that the Hessian matrix $\nabla^2 \ell^0$ (or equivalently, the information matrix) is non-singular.

In general, the non-singularity of the information matrix is a necessary and sufficient condition for local identification in a likelihood model where the vector of parameters has finite dimension (Rothenberg, 1971). Therefore, the non-singularity of the Hessian of the likelihood function is a necessary and sufficient condition for local identification in step-1. A necessary condition for local identification is also necessary for global identification. Thus, the sufficient conditions for global step-1 identification in Proposition 1 imply that the Hessian is non-singular.

5.1.2 Rank condition for step 2 identification

Given a label assignment, define the system of restrictions in step 2 as $\mathbf{c}(\boldsymbol{\pi}, \mathbf{P}) = 0$, where $\mathbf{c}(\boldsymbol{\pi}, \mathbf{P}) = \{c_i(a_i, \mathbf{x}, \kappa; \boldsymbol{\pi}, \mathbf{P}) : i \in \mathcal{I},\ a_i \in \mathcal{A} - \{0\},\ \kappa = 1, 2, ..., L_\kappa(\mathbf{x}),\ \mathbf{x} \in \mathcal{X}\}$ and:

$$c_i(a_i, \mathbf{x}, \kappa; \boldsymbol{\pi}, \mathbf{P}) \equiv \sum_{\boldsymbol{a}_{-i}} \left(\prod_{j \neq i} P_j(a_j | \mathbf{x}, \kappa) \right) \pi_i(a_i, \boldsymbol{a}_{-i}, \mathbf{x}, \kappa) - \pi_i^{\mathbf{P}}(a_i, \mathbf{x}, \kappa) \tag{22}$$

The number of restrictions in $\mathbf{c}(\boldsymbol{\pi}, \mathbf{P})$ is equal to the number of free probabilities in $\mathbf{P}$. We use $\nabla \mathbf{c}^0$ to represent the Jacobian matrix of $\mathbf{c}(\boldsymbol{\pi}, \mathbf{P})$ evaluated at the true $\left(\boldsymbol{\pi}^0, \mathbf{P}^0\right)$. We use $\nabla_{\boldsymbol{\pi}'} \mathbf{c}^0$ and $\nabla_{\mathbf{P}'} \mathbf{c}^0$ to represent the columns of this Jacobian associated to $\boldsymbol{\pi}$ and $\mathbf{P}$, respectively, such that $\nabla \mathbf{c}^0 \equiv [\nabla_{\boldsymbol{\pi}'} \mathbf{c}^0,\ \nabla_{\mathbf{P}'} \mathbf{c}^0]$. Note that the vector of functions $\mathbf{c}(\boldsymbol{\pi}, \mathbf{P})$ is linear in $\boldsymbol{\pi}$. We can represent $\mathbf{c}(\boldsymbol{\pi}, \mathbf{P})$ as $\mathbf{A}(\mathbf{P})\ \boldsymbol{\pi} - \mathbf{b}(\mathbf{P})$ where $\mathbf{b}(\mathbf{P})$ is a vector with elements $\{\pi_i^{\mathbf{P}}(a_i, \mathbf{x}, \kappa)\}$, and $\mathbf{A}(\mathbf{P})$ is a matrix with elements $\prod_{j \neq i} P_j(a_j | \mathbf{x}, \kappa)$ and zeroes. Therefore, the Jacobian matrix $\nabla_{\boldsymbol{\pi}'} \mathbf{c}^0$ is $\mathbf{A}^0 \equiv \mathbf{A}\left(\mathbf{P}^0\right)$.

As defined in section 4, $\boldsymbol{\pi}^0$ is globally identified in step 2 if: (a) there exists at least one label assignment for which $\boldsymbol{\pi}^0$ is the unique value of $\boldsymbol{\pi}$ that solves the system $\mathbf{c}\left(\boldsymbol{\pi}, \mathbf{P}^0\right) = \mathbf{0}$; and (b) there is no other label assignment delivering a different unique solution for $\boldsymbol{\pi}$. The rank condition for local identification in step 2 is that the Jacobian matrix $\nabla_{\boldsymbol{\pi}'} \mathbf{c}^0$ (or $\mathbf{A}^0$) is full column rank given at least one correct label assignment. This condition is equivalent to condition (a) in Proposition 3.

In summary, the rank condition for the sequential identification of $(\mathbf{h}^0, \mathbf{P}^0, \boldsymbol{\pi}^0)$ using steps 1

and 2 can be described in terms of the full-column rank of the matrix:

$$\mathbf{J}_{\text{seq}} \equiv \begin{bmatrix} \nabla^2_{\mathbf{hh}'}\ell^0 & \nabla^2_{\mathbf{hP}'}\ell^0 & \mathbf{0} \\ \nabla^2_{\mathbf{Ph}'}\ell^0 & \nabla^2_{\mathbf{PP}'}\ell^0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}^0 \end{bmatrix} \tag{23}$$

where we use $\nabla^2_{\mathbf{hh}'}\ell^0$, $\nabla^2_{\mathbf{hP}'}\ell^0$, $\nabla^2_{\mathbf{Ph}'}\ell^0$, and $\nabla^2_{\mathbf{PP}'}\ell^0$ to represent the submatrices that form the Hessian matrix.

5.2 Rank condition for joint identification

The true value $(\mathbf{h}^0, \mathbf{P}^0, \boldsymbol{\pi}^0)$ maximizes the likelihood $\ell(\mathbf{h}, \mathbf{P})$ subject to the constraints $\mathbf{c}(\boldsymbol{\pi}, \mathbf{P}) = 0$. The Lagrange function of this constrained maximum likelihood problem is $\mathcal{L}(\boldsymbol{\theta}) = \ell(\mathbf{h}, \mathbf{P}) + \boldsymbol{\lambda}'$ $\mathbf{c}(\boldsymbol{\pi}, \mathbf{P})$ where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers, and $\boldsymbol{\theta} \equiv (\mathbf{h}, \mathbf{P}, \boldsymbol{\pi}, \boldsymbol{\lambda})$. We say that the true value $\boldsymbol{\theta}^0 \equiv (\mathbf{h}^0, \mathbf{P}^0, \boldsymbol{\pi}^0, \boldsymbol{\lambda}^0)$ is point-identified if it is the unique solution of this constrained maximum likelihood problem. This Lagrange function $\mathcal{L}(\boldsymbol{\theta})$ is twice continuously differentiable in $\boldsymbol{\theta}$. Suppose that $\boldsymbol{\theta}^0$ is an interior point in the parameter space. Then, $\boldsymbol{\theta}^0$ should satisfy the Lagrange first order conditions of optimality.

$$\begin{aligned} \nabla_{\mathbf{h}}\ell^0 &= \mathbf{0} \\ \nabla_{\mathbf{P}}\ell^0 + [\nabla_{\mathbf{P}}\mathbf{c}^0]'\ \boldsymbol{\lambda}^0 &= \mathbf{0} \\ \mathbf{A}^{0\prime}\ \boldsymbol{\lambda}^0 &= \mathbf{0} \\ \mathbf{A}^0\ \boldsymbol{\pi}^0 - \mathbf{b}^0 &= \mathbf{0} \end{aligned} \tag{24}$$

where the last two sets conditions take into account the particular structure of the constraints $\mathbf{c}(\boldsymbol{\pi}, \mathbf{P}) = \mathbf{0}$.

In this constrained optimization problem, the true value of the vector of Lagrange multipliers, $\boldsymbol{\lambda}^0$, is zero. Remember that the Lagrange multipliers are the shadow prices of the constraints. By the information inequality, the true $(\mathbf{h}^0, \mathbf{P}^0)$ maximizes the likelihood $\ell(\mathbf{h}, \mathbf{P})$ (and satisfies the constraints) and there is not any value $(\mathbf{h}, \mathbf{P})$ that does not satisfy the constraints and implies a larger value for the likelihood. Therefore, the value of the shadow prices of the constraints, $\boldsymbol{\lambda}^0$ at any solution to the constrained problem should be zero. This implies that $(\mathbf{h}^0, \mathbf{P}^0, \boldsymbol{\pi}^0)$ should satisfy the following conditions:

$$\begin{aligned} \nabla_{\mathbf{h}}\ell^0 &= \mathbf{0} \\ \nabla_{\mathbf{P}}\ell^0 &= \mathbf{0} \\ \mathbf{A}^0\ \boldsymbol{\pi}^0 - \mathbf{b}^0 &= \mathbf{0} \end{aligned} \tag{25}$$

The rank condition for local joint identification is that the Jacobian matrix of this system of equations with respect to $(\mathbf{h}, \mathbf{P}, \boldsymbol{\pi})$ (and evaluated at $(\mathbf{h}^0, \mathbf{P}^0, \boldsymbol{\pi}^0)$) is full-column rank. This Jacobian

matrix has the following form:

$$\mathbf{J}_{joint} = \begin{bmatrix} \nabla^2_{\mathbf{hh'}}\ell^0 & \nabla^2_{\mathbf{hP'}}\ell^0 & \mathbf{0} \\ \nabla^2_{\mathbf{Ph'}}\ell^0 & \nabla^2_{\mathbf{PP'}}\ell^0 & \mathbf{0} \\ \mathbf{0} & \nabla_{\mathbf{P'}}\mathbf{c}^0 & \mathbf{A}^0 \end{bmatrix} \tag{26}$$

5.3 Relationship between sequential and joint identification

We can determine the relationship between sequential and joint identification by comparing the conditions for full-column rank of matrices $\mathbf{J}_{\text{seq}}$ and $\mathbf{J}_{joint}$. Note that the only difference between these two matrices is in the submatrix $\nabla_{\mathbf{P'}}\mathbf{c}^0$ that appears in $\mathbf{J}_{joint}$ but not in $\mathbf{J}_{\text{seq}}$. That is, joint identification rank conditions incorporate additional constraints on choice probabilities and parameters which are implied by equilibrium behavior and the structure of the model and may resolve step-1 under-identification.

Proposition 6 below presents necessary and sufficient conditions to have local joint identification without local step-1 identification. Before we present this Proposition, it is worthwhile to present several results on the relationship between joint and sequential identification that are straightforward implications of the full-column rank conditions of $\mathbf{J}_{\text{seq}}$ and $\mathbf{J}_{joint}$.[17]

(A) The condition of full-column rank of the matrices $\mathbf{A}^0$ *and* $\begin{bmatrix} \nabla^2_{\mathbf{hh'}}\ell^0 \\ \nabla^2_{\mathbf{Ph'}}\ell^0 \end{bmatrix}$ *is necessary both for sequential and for joint identification.* This follows from the structure of the columns in matrices $\mathbf{J}_{\text{seq}}$ and $\mathbf{J}_{joint}$. The condition of full column rank of matrix $\begin{bmatrix} \nabla^2_{\mathbf{hh'}}\ell^0 \\ \nabla^2_{\mathbf{Ph'}}\ell^0 \end{bmatrix}$ means that the mixture distribution $\mathbf{h}^0$ is locally identified if we know the vector of choice probabilities $\mathbf{P}^0$. Similarly, full column rank of matrix $\mathbf{A}^0$ is equivalent to say that the vector of payoffs $\boldsymbol{\pi}^0$ is identified if we know the vector of choice probabilities $\mathbf{P}^0$. That is, joint identification cannot help if failure of sequential identification occurs because mixture weights and/or payoffs are not identified given $\mathbf{P}^0$.

(B) Sequential identification implies joint identification. Sequential identification requires: (i) the Hessian matrix $\nabla^2\ell^0$ is full-column rank; and (ii) $\mathbf{A}^0$ is full-column-rank. Condition (i) implies that the matrix that results from vertically stacking the Hessian and $[\mathbf{0}\ ,\ \nabla_{\mathbf{P'}}\mathbf{c}^0]$ is also full-column rank, regardless the value of $\nabla_{\mathbf{P'}}\mathbf{c}^0$. Then, full column rank of $\mathbf{J}_{joint}$ follows by contradiction.

(C) Joint identification implies step 2 identification. Joint identification requires $\mathbf{A}^0$ to be full-column-rank, and this implies step-2 identification.

[17] We are comparing rank conditions at a correct label assignment only. Note that the only effect of label reassignments on the rank of matrices $\mathbf{J}_{\text{seq}}$ and $\mathbf{J}_{joint}$ is through the blocks of matrix $\mathbf{A}^0$. Extended rank conditions which remove incorrect label assignments, such as the ones in condition (b) of Proposition 3, are necessary and sufficient for both joint and sequential global identification.

(D) LEMMA 2. The Jacobian matrix $\nabla_{\mathbf{P}'}\mathbf{c}^0$ is non-singular if the equilibria in the DGP are regular.[18] ■

Proof: By an appropriate rearrangement of the rows in the system of equations $\mathbf{c}(\boldsymbol{\pi}, \mathbf{P}) = 0$, we can represent it as a set of sub-systems one for each value of $(\mathbf{x}, \kappa)$. Let $\mathbf{c}\left(\mathbf{x}, \kappa, \boldsymbol{\pi}_{(\mathbf{x},\kappa)}, \mathbf{P}_{(\mathbf{x},\kappa)}\right) = 0$ be the sub-system for a value of $(\mathbf{x}, \kappa)$, where $\boldsymbol{\pi}_{(\mathbf{x},\kappa)}$ and $\mathbf{P}_{(\mathbf{x},\kappa)}$ represent the vectors of pay-offs and choice probabilities associated to $(\mathbf{x}, \kappa)$. Since the vector of choice probabilities $\mathbf{P}_{(\mathbf{x},\kappa)}$ enters *only* in the sub-system of equations for the corresponding value $(\mathbf{x}, \kappa)$, we have that the Jacobian matrix $\nabla_{\mathbf{P}'}\mathbf{c}^0$ is a block diagonal matrix that contains the Jacobians $\nabla_{\mathbf{P}'}\mathbf{c}^0_{(\mathbf{x},\kappa)}$ for each value of $(\mathbf{x}, \kappa)$. Therefore, $\nabla_{\mathbf{P}'}\mathbf{c}^0$ is non-singular if and only if the Jacobians $\nabla_{\mathbf{P}'}\mathbf{c}^0_{(\mathbf{x},\kappa)}$ are non-singular for each value of $(\mathbf{x}, \kappa)$. We can represent the equilibrium equations of the model as $\mathbf{P}_{(\mathbf{x},\kappa)} - \boldsymbol{\Psi}(\mathbf{x}, \kappa, \mathbf{P}_{(\mathbf{x},\kappa)}) = 0$. The definition of regular equilibrium is that the Jacobian matrix $\partial[\mathbf{P}_{(\mathbf{x},\kappa)} - \boldsymbol{\Psi}(\mathbf{x}, \kappa, \mathbf{P}_{(\mathbf{x},\kappa)})]/\partial \mathbf{P}'_{(\mathbf{x},\kappa)}$ is non-singular. By definition, the system $\mathbf{c}\left(\mathbf{x}, \kappa, \boldsymbol{\pi}_{(\mathbf{x},\kappa)}, \mathbf{P}_{(\mathbf{x},\kappa)}\right)$ is equal to $\Lambda^{-1}[\mathbf{P}_{(\mathbf{x},\kappa)}] - \Lambda^{-1}[\boldsymbol{\Psi}(\mathbf{x}, \kappa, \mathbf{P}_{(\mathbf{x},\kappa)})]$, where $\Lambda[.]$ is the function mapping choice-specific utilities into choice probabilities. This well-known mapping is invertible so the Jacobian matrix Λ_P^{-1} is non-singular. Therefore, we have that $\nabla_{\mathbf{P}'}\mathbf{c}^0_{(\mathbf{x},\kappa)} = \Lambda_P^{-1}\, \partial[\mathbf{P}_{(\mathbf{x},\kappa)} - \boldsymbol{\Psi}(\mathbf{x}, \kappa, \mathbf{P}_{(\mathbf{x},\kappa)})]/\partial \mathbf{P}'_{(\mathbf{x},\kappa)}$, so $\nabla_{\mathbf{P}'}\mathbf{c}^0_{(\mathbf{x},\kappa)}$ is non-singular because the product of non-singular matrices is non-singular. ■

PROPOSITION 6. We have local joint identification without local sequential identification if and only if the following conditions hold: (i) The Hessian matrix $\nabla^2\ell^0$ is singular, but its submatrix $\begin{bmatrix} \nabla^2_{\mathbf{hh}'}\ell^0 \\ \nabla^2_{\mathbf{Ph}'}\ell^0 \end{bmatrix}$ is full-column rank; (ii) matrix $\mathbf{A}^0$ is full-column rank; and (iii) the column space (or span) of matrix $\mathbf{A}^0$ does not include any of the vectors $\nabla_{\mathbf{P}'}\mathbf{c}^0\, \boldsymbol{\beta}_P$, where $\boldsymbol{\beta}_P$ is the $\mathbf{P}$-component of any vector $(\boldsymbol{\beta}'_h, \boldsymbol{\beta}'_P)$ from the nullspace of $\nabla^2\ell^0$. ■

Proof: Suppose that conditions *(i)* and *(ii)* hold. We need to show that it is possible to have a matrix $\mathbf{J}_{joint}$ that is full-column rank so there is joint identification. First, since $\nabla^2\ell^0$ is not full-column rank, there should be vectors $\boldsymbol{\beta}_h$ and $\boldsymbol{\beta}_P$ with $(\boldsymbol{\beta}'_h, \boldsymbol{\beta}'_P) \neq \mathbf{0}$ such that:

$$\begin{bmatrix} \nabla^2_{\mathbf{hh}'}\ell^0 \\ \nabla^2_{\mathbf{Ph}'}\ell^0 \end{bmatrix} \boldsymbol{\beta}_h + \begin{bmatrix} \nabla^2_{\mathbf{hP}'}\ell^0 \\ \nabla^2_{\mathbf{PP}'}\ell^0 \end{bmatrix} \boldsymbol{\beta}_P = \mathbf{0} \tag{27}$$

Recall that the nullspace of the Hessian matrix, $Null(\nabla^2\ell^0)$, is defined as the set of vectors $(\boldsymbol{\beta}'_h, \boldsymbol{\beta}'_P) \neq \mathbf{0}$ that satisfy equation (27). Note that any vector in this nullspace cannot have $\boldsymbol{\beta}_h \neq \mathbf{0}$ and $\boldsymbol{\beta}_P = \mathbf{0}$ because this contradicts the condition that $\begin{bmatrix} \nabla^2_{\mathbf{hh}'}\ell^0 \\ \nabla^2_{\mathbf{Ph}'}\ell^0 \end{bmatrix}$ is full-column rank.

[18] Note that Assumption 3(C) establishes that all the equilibria in the DGP are regular.

Therefore, all vectors in this nullspace should have $\boldsymbol{\beta}_P \neq \mathbf{0}$. Since $\boldsymbol{\beta}_P \neq \mathbf{0}$ and $\nabla_{\mathbf{P}'}\mathbf{c}^0$ is non-singular at a regular equilibrium, it should be true that $\nabla_{\mathbf{P}'}\mathbf{c}^0\ \boldsymbol{\beta}_P \neq \mathbf{0}$, otherwise we would find the contradiction that $\boldsymbol{\beta}_P = 0$. Given these conditions,we have that matrix $\mathbf{J}_{joint}$ is full-column rank if and only if, for any value of $\boldsymbol{\beta}_P$ in $Null(\nabla^2\ell^0)$, the vector $\nabla_{\mathbf{P}'}\mathbf{c}^0\ \boldsymbol{\beta}_P$ does not belong to the column space (span) of matrix $\mathbf{A}^0$. ■

Conditions (i) and (ii) in Proposition 6 are quite intuitive identification conditions as discussed in result (A) above. Condition (iii) has a clear interpretation in term a matrix algebra but it is not intuitive from the point of view of identification. The following corollaries of Proposition 6 are intuitive conditions which are necessary for condition (iii) in Proposition 6.

COROLLARY 1. A necessary condition for condition (iii) is that the number of restrictions $\mathbf{c}(\boldsymbol{\pi}, \mathbf{P}) = 0$ is strictly greater than the number of parameters in $\boldsymbol{\pi}$. Otherwise, $\mathbf{A}^0$ is a square matrix, and given that it is non-singular (condition (ii)) its column space includes any vector $\nabla_{\mathbf{P}'}\mathbf{c}^0$ $\boldsymbol{\beta}_P$. Therefore, if the number of restrictions is equal to the number of payoff parameters, then $\mathbf{J}_{joint}$ is full-column rank if and only if $\mathbf{J}_{\text{seq}}$ is full-column rank.

COROLLARY 2. A necessary condition for condition (iii) is that the number of free probabilities in the distribution $Q(\boldsymbol{a}|\mathbf{x})$ is greater or equal than the number of structural parameters $(\mathbf{h}, \boldsymbol{\pi})$.

Ideally, we would like that the identification conditions in Proposition 6 were conditions on the primitives of the model. With that type of sufficient conditions, we would be sure that there is a subclass of DGPs where we have local joint identification without local sequential identification. In the absence of sufficient conditions on the primitives of the model, it is reasonable to wonder whether there are DGPs within our class of models that satisfy the necessary and sufficient conditions in Proposition 6. Unfortunately, deriving this type of sufficient conditions on the primitives of the model is a very complicated task. Instead, in the following section we study a general class of games where sequential identification never applies. We particularize the conditions of Proposition 6 to this example, and provide more intuitive identification conditions for this case. We also present a numerical example showing that there is a continuum of parameter values (of DGPs) for which there is not sequential identification but there is local joint identification.

5.4 Example: 2 x 2 x 2 games

Consider a game with two players, binary choice, and two points in the support of the unobserved market heterogeneity κ. Let $\{A, B\}$ be the two values of the unobserved type. We use $h_A(\mathbf{x})$ to represent $h(A|\mathbf{x})$, and $P_{iA}(\mathbf{x})$ and $P_{iB}(\mathbf{x})$ to represent the CCPs $P_i(a_i = 1|\mathbf{x}, \kappa = A)$ and $P_i(a_i = 1|\mathbf{x}, \kappa = B)$, respectively. The vector $\mathbf{x}$ has two variables, $z_1 \in \mathcal{Z}$ and $z_2 \in \mathcal{Z}$, with $\mathcal{Z} \equiv \{z^{(1)}, z^{(2)}, ..., z^{(|\mathcal{Z}|)}\}$. We use $\pi_{iA}(a_j, z_i)$ and $\pi_{iB}(a_j, z_i)$ to represent the payoff of player i when $a_i = 1$ and for $\kappa = A$ and $\kappa = B$, respectively.

5.4.1 Proposition 6 in 2 x 2 x 2 games

First, we particularize to this model conditions (i) to (iii) in Proposition 6.

Condition (i). "The Hessian matrix $\nabla^2 \ell^0$ is always singular". Hall and Zhou (2003) have proved this result for a general nonparametric finite mixture model with two independent discrete random variables (e.g., two players) and two points of support for the unobserved heterogeneity. We briefly illustrate this result for our 2 x 2 x 2 game. Let $\ell^x(\mathbf{x}; h_A(\mathbf{x}), \mathbf{P}(\mathbf{x}))$ be the population log-likelihood conditional on a value of $\mathbf{x}$, where $\mathbf{P}(\mathbf{x})$ is the vector of CCPs $(P_{1A}(\mathbf{x}),\ P_{1B}(\mathbf{x}),\ P_{2A}(\mathbf{x}),\ P_{2B}(\mathbf{x}))'$. This log-likelihood function is:

$$\ell^x(\mathbf{x}; h_A(\mathbf{x}), \mathbf{P}(\mathbf{x})) = \sum_{(a_1,a_2)\in\{0,1\}^2} Q(a_1, a_2|\mathbf{x}) \ln \left[\begin{array}{c} h_A(\mathbf{x})\ d(a_1, P_{1A}(\mathbf{x}))\ d(a_2, P_{2A}(\mathbf{x})) \\ +(1 - h_A(\mathbf{x}))\ d(a_1, P_{1B}(\mathbf{x}))\ d(a_2, P_{2B}(\mathbf{x})) \end{array} \right] \tag{28}$$

with $d(a, P) \equiv P^a\ (1-P)^{1-a}$. The full log-likelihood function is $\ell(\mathbf{h}, \mathbf{P}) = \sum_{\mathbf{x}\in\mathcal{X}} p(\mathbf{x})\ \ell^x(\mathbf{x}; h_A(\mathbf{x}), \mathbf{P}(\mathbf{x}))$. Note that the vector of parameters $(h_A(\mathbf{x}), \mathbf{P}(\mathbf{x}))$ associated to $\mathbf{x}$ enters only in the log-likelihood $\ell^x(\mathbf{x}; h_A(\mathbf{x}), \mathbf{P}(\mathbf{x}))$ and not in the log-likelihood for any other value of $\mathbf{x}$. Therefore, the Hessian matrix of the complete log-likelihood function is a block diagonal matrix of the $\mathbf{x}-$specific Hessians $\nabla^2 \ell^x(\mathbf{x})$, i.e., $\nabla^2 \ell = diag\left\{\nabla^2 \ell^x(\mathbf{x}) :\ \text{for } \mathbf{x} \in \mathcal{X}\right\}$. This implies that the Hessian matrix $\nabla^2 \ell^0$ is non-singular if and only if the Hessian matrices $\nabla^2 \ell^{x,0}(\mathbf{x})$ are non-singular for every value of $\mathbf{x}$. The Hessian evaluated at true parameter values is equal to minus the variance-covariance matrix of the vector of scores. The variance of a random vector is singular if and only if the components of the vector are linearly dependent with probability one. Linear dependence always follows if the support of the random vector has fewer points than the number of elements in the vector. In this model, the score vector has 5 components and its support is (at most) the number of points in the support of (a_1, a_2) which is 4. Therefore, the Hessian matrix is always singular in this model.

Condition (i). "Submatrix $\nabla^2_{[\mathbf{h},\mathbf{P}]\mathbf{h}'}\ell^0$ is full-column rank". As described above, the Hessian matrix has a block diagonal structure. Therefore, submatrix $\nabla^2_{[\mathbf{h},\mathbf{P}]\mathbf{h}'}\ell^0$ is equal to $diag\{\nabla^2_{[h,\mathbf{P}]h}\ell^{x,0}(\mathbf{x}) : \text{for } \mathbf{x} \in \mathcal{X}\}$, where $\nabla^2_{[h,\mathbf{P}]h}\ell^{x,0}(\mathbf{x})$ is the 5×1 vector $(\nabla^2_{hh}\ell^{x,0},\ \nabla^2_{hP_{1A}}\ell^{x,0},\ \nabla^2_{hP_{1B}}\ell^{x,0}\ \nabla^2_{hP_{2A}}\ell^{x,0},\ \nabla^2_{hP_{2B}}\ell^{x,0})'$. The condition that a vector has full column rank is that at least one of its elements is not zero. We show in the Appendix that $\nabla^2_{[h,\mathbf{P}]h}\ell^{x,0}$ is full column rank if and only if $P_{iA} \neq P_{iB}$ for at least one of the players. Therefore, we need that for every value of $\mathbf{x}$ the unobserved heterogeneity has an effect on the choice probability of at least one player. Otherwise, it is clear that the mixture weight h cannot be identified.

Condition (ii). "Matrix $\mathbf{A}^0$ is full-column rank". For every player i, unobserved type κ, and state variables $(z_1, z_2) \in \mathcal{Z}^2$, we have a constraint $c_{i\kappa}(z_1, z_2; \boldsymbol{\pi}, \mathbf{P}) = 0$ with:

$$c_{i\kappa}(z_1, z_2; \boldsymbol{\pi}, \mathbf{P}) \equiv [1 - P_{j\kappa}(z_1, z_2)]\ \pi_{i\kappa}(0, z_i) + P_{j\kappa}(z_1, z_2)\ \pi_{i\kappa}(1, z_i) - \Lambda^{-1}\left(P_{i\kappa}(z_1, z_2)\right) \tag{29}$$

For a value of (i, κ, z_i), we can define the $|\mathcal{Z}| \times 1$ vector of restrictions $\mathbf{c}_{i\kappa}(z_i; \boldsymbol{\pi}, \mathbf{P})$ for every value of z_j with $j \neq i$. Note that the vector $\mathbf{c}_{i\kappa}(z_i; \boldsymbol{\pi}, \mathbf{P})$ depends on $\boldsymbol{\pi}$ only through the 2×1 vector of payoffs $\boldsymbol{\pi}_{i\kappa}(z_i) \equiv (\pi_{i\kappa}(0, z_i),\ \pi_{i\kappa}(1, z_i))'$. Furthermore, the payoff vector $\boldsymbol{\pi}_{i\kappa}(z_i)$ only enters in the restrictions $\mathbf{c}_{i\kappa}(z_i; \boldsymbol{\pi}, \mathbf{P})$ and not in restrictions for values different to (i, κ, z_i). Therefore, the Jacobian matrix $\mathbf{A}^0 \equiv \nabla_{\pi'}\mathbf{c}^0$ has a block-diagonal structure, i.e., $\mathbf{A}^0 = diag\left\{\mathbf{A}^0_{i\kappa}(z_i) : \text{for any } (i, \kappa, z_i)\right\}$ where $\mathbf{A}^0_{i\kappa}(z_i) \equiv \nabla_{\boldsymbol{\pi}_{i\kappa}(z_i)'}\mathbf{c}^0_{i\kappa}(z_i)$. This implies that $\mathbf{A}^0$ is full column rank if and only if every submatrix $\mathbf{A}^0_{i\kappa}(z_i)$ is full column rank. Taking into account the structure of the constraints in equation (29), we have that:

$$\mathbf{A}^0_{i\kappa}(z_i) = \begin{bmatrix} 1 - P_{j\kappa}(z_i, z_j^{(1)}), & P_{j\kappa}(z_i, z_j^{(1)}) \\ 1 - P_{j\kappa}(z_i, z_j^{(2)}), & P_{j\kappa}(z_i, z_j^{(2)}) \\ \vdots & \vdots \\ 1 - P_{j\kappa}(z_i, z_j^{(|\mathcal{Z}|)}), & P_{j\kappa}(z_i, z_j^{(|\mathcal{Z}|)}) \end{bmatrix} \tag{30}$$

A necessary and sufficient condition for $\mathbf{A}^0_{i\kappa}(z_i)$ to be full column rank is that there are two values of z_j, say $z_j^{(a)}$ and $z_j^{(b)}$, such that $P_{j\kappa}(z_i, z_j^{(a)}) \neq P_{j\kappa}(z_i, z_j^{(b)})$. Therefore, $\mathbf{A}^0$ is full column rank if this condition holds for every value of (i, κ, z_i). That is, for any value of (i, κ, z_i) the probability $P_{j\kappa}(z_i, z_j)$ varies with z_j such that the exclusion restriction has power.

Condition (iii). "The column space (or span) of matrix $\mathbf{A}^0$ does not include any of the vectors $\nabla_{\mathbf{P}'}\mathbf{c}^0\ \boldsymbol{\beta}_P$". We consider here the three necessary conditions established in the Corollaries 1 and 2.

Corollary 1: "The number of restrictions $\mathbf{c}(\boldsymbol{\pi}, \mathbf{P}) = 0$ is strictly greater than the number of parameters in $\boldsymbol{\pi}$". In this example, the number of restrictions is $4\ |\mathcal{Z}|^2$ and the number of payoff

parameters is 8 $|\mathcal{Z}|$. Therefore, condition (iii) requires the support set $\mathcal{Z}$ to have 3 points or more: $|\mathcal{Z}| \geq 3$.

Corollary 2: "The number of free probabilities in the distribution $Q(\boldsymbol{a}|\mathbf{x})$ is greater or equal than the number of structural parameters $(\mathbf{h}, \boldsymbol{\pi})$." In this example, the number of free probabilities in Q is $3|\mathcal{Z}|^2$, and the number of structural parameters in $(\mathbf{h}, \boldsymbol{\pi})$ is $|\mathcal{Z}|^2 + 8\ |\mathcal{Z}|$. Therefore, this order condition requires $|\mathcal{Z}| \geq 4$. Note that this order condition is tighter that the one that comes from Corollary 1.

Non-singularity of the square Jacobian matrix $\nabla_{\mathbf{P}}\mathbf{c}^0$. Let $\mathbf{c}_{(z_1,z_2,\kappa)}(\boldsymbol{\pi}, \mathbf{P})$ be the vector of functions $c_{i\kappa}(z_1, z_2; \boldsymbol{\pi}, \mathbf{P})$ associated to a particular value of (z_1, z_2, κ), i.e., $\mathbf{c}_{(z_1,z_2,\kappa)}(\boldsymbol{\pi}, \mathbf{P})$ is the 2×1 vector $(c_{1\kappa}(z_1, z_2; \boldsymbol{\pi}, \mathbf{P}),\ c_{2\kappa}(z_1, z_2; \boldsymbol{\pi}, \mathbf{P}))'$. Given the expression of $c_{i\kappa}(z_1, z_2; \boldsymbol{\pi}, \mathbf{P})$ in equation (29), we have that $\mathbf{c}_{(z_1,z_2,\kappa)}(\boldsymbol{\pi}, \mathbf{P})$ depends on $\mathbf{P}$ only through the two probabilities $\mathbf{P}_{(z_1,z_2,\kappa)} = (P_{1\kappa}(z_1, z_2),\ P_{2\kappa}(z_1, z_2))$. Furthermore, these probabilities do not enter in the vector of constraints $\mathbf{c}_{(z_1',z_2',\kappa')}(\boldsymbol{\pi}, \mathbf{P})$ for values (z_1', z_2', κ') different to (z_1, z_2, κ). Therefore, the Jacobian matrix $\nabla_{\mathbf{P}'}\mathbf{c}^0$ has a block diagonal structure: $\nabla_{\mathbf{P}'}\mathbf{c}^0 = diag\{\nabla_{\mathbf{P}_{(z_1,z_2,\kappa)}}\mathbf{c}^0_{(z_1,z_2,\kappa)} : \text{for any value } (z_1, z_2, \kappa)\}$. The Jacobian matrix $\nabla_{\mathbf{P}'}\mathbf{c}^0$ is non-singular if and only if every (sub) Jacobian matrix $\nabla_{\mathbf{P}_{(z_1,z_2,\kappa)}}\mathbf{c}^0_{(z_1,z_2,\kappa)}$ is non-singular. Given equation (29), we have that this Jacobian is the following 2×2 matrix:

$$\nabla_{\mathbf{P}_{(z_1,z_2,\kappa)}}\mathbf{c}^0_{(z_1,z_2,\kappa)} = \begin{bmatrix} \varphi(P_{1\kappa}(z_1, z_2)), & \pi_{1\kappa}(1, z_1) - \pi_{1\kappa}(0, z_1) \\ \pi_{2\kappa}(1, z_2) - \pi_{2\kappa}(0, z_2), & \varphi(P_{1\kappa}(z_1, z_2)) \end{bmatrix} \tag{31}$$

where $\varphi(.)$ is the derivative of the Quantile function (inverse CDF) of the private information variable ε.[19] It is clear that this matrix is singular only if $\varphi(P_{1\kappa}(z_1, z_2))\ \varphi(P_{2\kappa}(z_1, z_2)) = [\pi_{1\kappa}(1, z_1) - \pi_{1\kappa}(0, z_1)]\ [\pi_{2\kappa}(1, z_2) - \pi_{2\kappa}(0, z_2)]$. This condition corresponds to a an equilibrium that is a singularity point, as described in Definition 2. The condition that the equilibrium is regular implies that this matrix is non-singular.

Unfortunately, even for this relatively simple model, the complete characterization of condition (iii), in Proposition 6, is not much simpler or intuitive than for the general model. Therefore, we present here a specific numerical example.

5.4.2 Numerical example

The following numerical example shows that there is a continuum of DGPs for the 2 x 2 x 2 game where there is joint identification of the structural parameters but there is not sequential

[19] For instance, if $\Lambda(.)$ is the logistic function (Logit model), we have that $\Lambda^{-1}(p) = \ln(p) - \ln(1-p)$ and $\varphi(p) = 1/p - 1/(1-p)$.

identification.

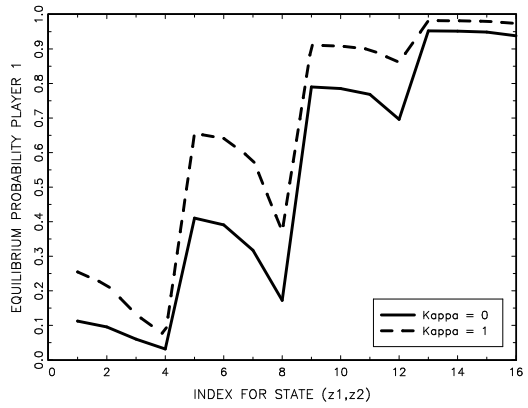
Description of DGPs. The distribution of the private information is Logistic. The support $\mathcal{Z}$ of the exogenous variables z_i consists of $|\mathcal{Z}|$ points uniformly spaced over the interval $[0,1]$ including the values 0 and 1. That is, $\mathcal{Z} = \{0, \frac{1}{|\mathcal{Z}|-1}, \frac{2}{|\mathcal{Z}|-1}, ..., \frac{|\mathcal{Z}|-2}{|\mathcal{Z}|-1}, 1\}$. We implement numerical experiments for different values of $|\mathcal{Z}|$. The support of the unobserved heterogeneity κ is $\{A, B\} = \{0, 1\}$. In all our experiments, the unobserved heterogeneity is payoff relevant and there is not multiple equilibria in the DGP. For the results that we report here, the probability distribution $h_A(z_1, z_2) = \overline{h}_A$ for any value of (z_1, z_2), where $\overline{h}_A$ is a constant such that there is independence between κ and (z_1, z_2). Of course, this independence is not known by the researcher, who estimates all the parameters $h_A(z_1, z_2)$ without restrictions. We have also implemented numerical experiments where the values of the probabilities $h_A(z_1, z_2)$ are generated as random draws from a Uniform (0,1) distribution, and we have obtained the same identification results as the ones reported here. In the experiments that we report here, we consider different values for the parameter $\overline{h}_A$.

The form of the payoff functions is: $\pi_{i\kappa}(0, z_i) = \alpha_{i0} + \alpha_{iz}\ z_i + \alpha_{i\kappa}\ \kappa$ and $\pi_{i\kappa}(1, z_i) - \pi_{i\kappa}(0, z_i) = \beta_{i0} + \beta_{iz}\ z_i + \beta_{i\kappa}\ \kappa$. Again, this linear form of the payoff function is not known to the researcher, who estimates unrestricted payoff functions $\pi_{i\kappa}(a_j, z_i)$. For the results that we report here, the values for the coefficients α and β are: $\alpha_{10} = -2$; $\alpha_{1z} = 5$; $\alpha_{1\kappa} = 1$; $\beta_{10} = -2$; $\beta_{1z} = 0$; $\beta_{1\kappa} = 0$; $\alpha_{20} = -3$; $\alpha_{2z} = 4$; $\alpha_{2\kappa} = 0.5$; $\beta_{20} = -3$; $\beta_{2z} = 0$; and $\beta_{2\kappa} = 0$. That is, $\pi_{1\kappa}(0, z_1) = -2 + 5\ z_1 + \kappa$, $\pi_{1\kappa}(1, z_1) - \pi_{1\kappa}(0, z_1) = -2$, $\pi_{2\kappa}(0, z_2) = -3 + 4\ z_2 + 0.5\kappa$, and $\pi_{2\kappa}(1, z_2) - \pi_{2\kappa}(0, z_2) = -3$. We can interpret this DGP as a model of market entry because the strategic interactions $\pi_{i\kappa}(1, z_i) - \pi_{i\kappa}(0, z_i)$ are negative. As we describe below, we have also considered DGPs where payoff parameters are randomly chosen, and we have obtained the same identification results.

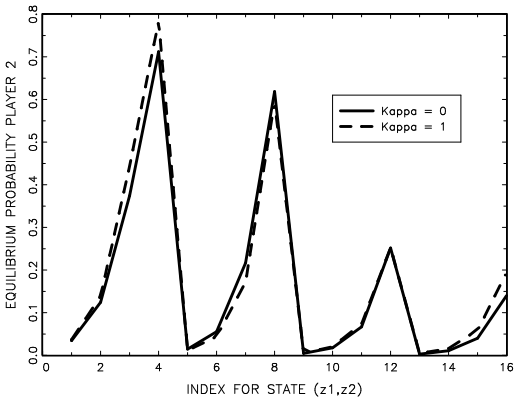
For every value of (z_1, z_2, κ), we compute an equilibrium by iterating simultaneously in the best response probability functions of the two players, and using as the initial value of this algorithm the vector of probabilities $(P_1, P_2) = (0, 0)$. We have always converged to an equilibrium using this procedure. Figure 3 presents the equilibrium probabilities for the two players for the DGP with $|\mathcal{Z}| = 4$.

Figure 3: Equilibrium Probabilities

Player 1



Player 2



Local identification.[20] Table 2 presents local identification results for DGPs with different values of $|\mathcal{Z}|$ and with $\overline{h}_A = 0.7$. For the Jacobian matrices J_{seq}, J_{joint}, $\mathbf{A}^0$, and $\nabla^2_{[\mathbf{h},\mathbf{P}]\mathbf{h}'}\ell^0$, we report their number of columns, their rank, and the minimum absolute eigenvalue of the corresponding square matrix, e.g., for the Jacobian matrix $\mathbf{A}^0$, the minimum absolute eigenvalue of the square matrix $\mathbf{A}^{0\prime}\mathbf{A}^0$. This minimum absolute eigenvalue is strictly greater than zero if and only if the original matrix is full column rank.[21]

In table 2, we can see that for any value of $|\mathcal{Z}|$ the matrices $\mathbf{A}^0$ and $\nabla^2_{[\mathbf{h},\mathbf{P}]\mathbf{h}'}\ell^0$ are full column rank such that conditions (i) and (ii) in Proposition 6 hold. Also, as we expect, for any value of $|\mathcal{Z}|$ the Jacobian matrix J_{seq} is not full column rank and there is not sequential identification. In fact, as the dimension of the state space increases, the degree of under-identification also increases, as measured by the difference between the number of columns and the rank of matrix J_{seq}. As shown above, the application of *Corollary 2* to this 2 x 2 x 2 model implies that a necessary condition for joint identification is that $|\mathcal{Z}| \geq 4$. This result is illustrated in table 2: when $|\mathcal{Z}| = 3$, the Jacobian matrix J_{joint} is not full column rank. However, for any value of $|\mathcal{Z}|$ greater or equal than four, this DGP has a Jacobian matrix J_{joint} that is full column rank such that there is local joint

[20] The code in GAUSS that we have used for the implementation of these numerical experiments can be found at `http://individual.utoronto.ca/vaguirre/wpapers/mequidata_endpaper_example.gss`

[21] The rank and the eivengavalues of these matrices are computed using the GAUSS commands `rank` and `eig`, respectively.

identification without sequential identification. Furthermore, for $|\mathcal{Z}| \geq 4$, the minimum absolute eigenvalue of $(J_{joint})'J_{joint}$ increases with $|\mathcal{Z}|$ such that identification becomes "stronger".[22]

Table 2
Numerical example of local joint identification
Two-players, Binary choice, Two unobserved types

DGP: Parameters $h_A(z_1, z_2)$ are equal to 0.7 for any value of (z_1, z_2)
Payoffs: $\pi_{i\kappa}(0, z_i) = \alpha_{i0} + \alpha_{iz}\ z_i + \alpha_{i\kappa}\ \kappa$,
and $\pi_{i\kappa}(1, z_i) - \pi_{i\kappa}(0, z_i) = \beta_{i0} + \beta_{iz}\ z_i + \beta_{1\kappa}\ \kappa$,
with $\alpha_{10} = -2$, $\alpha_{1z} = 5$, $\alpha_{1\kappa} = 1$, $\beta_{10} = -2$, $\beta_{1z} = 0$, $\beta_{1\kappa} = 0$,
$\alpha_{20} = -3$, $\alpha_{2z} = 4$, $\alpha_{2\kappa} = 0.5$, $\beta_{20} = -3$, $\beta_{2z} = 0$, $\beta_{2\kappa} = 0$

|Z|		matrix J_{seq}	matrix J_{joint}	matrix $\mathbf{A}^0$	matrix $\nabla^2_{[\mathbf{h},\mathbf{P}]\mathbf{h}'}\ell^0$
|Z|=3	columns	69	69	24	9
	rank	51	63	24	9
	min abs eig.	$8.7*10^{-18}$	$7.4*10^{-18}$	0.0111	0.4066
|Z|=4	columns	112	112	32	16
	rank	80	112	32	16
	min abs eig.	$5.6*10^{-18}$	$2.8*10^{-5}$	0.0119	0.4066
|Z|=12	columns	816	816	96	144
	rank	528	816	96	144
	min abs eig.	$6.6*10^{-18}$	$6.3*10^{-4}$	0.0211	0.3876
|Z|=20	columns	2160	2160	160	400
	rank	1360	2160	160	400
	min abs eig.	$1.4*10^{-18}$	$1.9*10^{-3}$	0.0313	0.3877

This numerical example illustrates that there are DGPs where the conditions of Proposition 6 hold. Still, it might be possible that there is only a finite number of DGPs where these conditions hold such that the conditions of Proposition 6 do not hold *generically*. To show that this is not the case, we have implemented two types of numerical experiments. First, we consider a set of 100 DGPs with the same parameter values as those in the DGP in table 2 expect for parameters $\overline{h}_A$ and

[22] For $|\mathcal{Z}| \geq 4$, the minimum absolute eigenvalue of matrix $(J_{joint})'J_{joint}$ is small. However, there is a substantial difference, of several orders of magnitude, between the computed minimum absolute eigenvalues of matrices $(J_{\text{seq}})'J_{\text{seq}}$ and $(J_{joint})'J_{joint}$. Similarly, there is the same order of magnitude in the difference between the minimum absolute eigenvalues of the matrices $(J_{joint})'J_{joint}$ when $|\mathcal{Z}| = 3$ and when $|\mathcal{Z}| \geq 4$. Therefore, there is a clear cut between identification and no-identification in this numerical example.

α_{10} that are changed marginally. Remember that in the DGP table 2 we have $(\overline{h}_A,\alpha_{1z}) = (0.7,5)$. Then, in this experiment we consider 100 DGPs where the values of $(\overline{h}_A,\alpha_{1z})$ are those in the grid $\{0.65, 0.66, 0.67, 0.68, 0.69, 0.71, 0.72, 0.73, 0.74, 0.75\} \times \{4.95, 4.96, 4.97, 4.98, 4.99, 5.01, 5.02, 5.03, 5.04, 5.05\}$. For all these DGPs we obtain the same identification results as in table 2. The second experiment consists of 100 DGPs where the values of the probabilities $h_A(z_1, z_2)$ are independent random draws from a $Uniform(0, 1)$ distribution, and the values of the payoffs $\pi_{i\kappa}(a_j, z_i)$ are independent random draws from a $Normal(0, 1)$ distribution. Again, for all these 100 "random" DGPs we obtain the same identification results as in table 2. We interpret these results as evidence that there is a continuum of DGPs where the conditions of Proposition 6 hold.

6 Conclusion

In empirical applications of games of incomplete information, we typically find that conditional on observable exogenous variables players' actions are correlated. One possible interpretation of this correlation is that common knowledge unobservables are present. Some of these unobservables may be payoff relevant while others may be 'sunspots' that affect players' beliefs and the selected equilibrium but do not have a direct effect on players' payoffs. This paper is motivated by the following question: is it possible to separate empirically the contribution of unobservables that affect the selection of an equilibrium in the data (i.e., non-payoff relevant unobservables or "sunspots") from the contribution of unobservables that are payoff-relevant? Is it possible to conclude that we need multiple equilibria to explain players' observed behavior?

We investigate this question by studying semiparametric identification of games when we allow for three types of unobserved heterogeneity for the researcher: payoff-relevant variables that are private information of each player (PI unobservables); payoff-relevant variables that are common knowledge to all the players (PR unobservables); and variables that are common knowledge to all the players, are not payoff-relevant but affect the equilibrium selection (multiple equilibria or ME unobservables). Two types of restrictions are crucial for our identification results: independence between private players' private information, and a exclusion restriction in the payoff function.

We show that implementation of a sequential identification/estimation approach requires that the researcher be able to match unobserved types across different values of the explanatory variables. We show that this problem of matching unobserved types can also appear in the sequential estimation of single-agent models. We provide necessary and sufficient conditions for the identifica-

tion of payoffs in step 2 under this matching-types problem. We also show that some restrictions on the primitives (e.g., additive separability of the unobservables in the payoff function) are sufficient conditions for identification.

Our results show that it is possible to separately identify the relative contributions of payoff-relevant and multiple equilibria unobserved heterogeneity to observed players' behavior. As De Paula and Tang (2012) and others have shown, multiplicity of equilibria can help identify some elements of the structure such as the sign of strategic interactions. However, without the exclusion restrictions that are needed to identify payoffs, it does not seem possible for the researcher to ascertain *ex-ante* that the correlation between the actions of players is induced by the occurrence of multiple equilibria in the data.

APPENDIX. Proof of Lemma 1

Proof. The proof of Lemma 1(A) is in the spirit of the proof in Mas-Colell et al. (1995; Propositions 17.D.3, 17.D.4, and 17.D.5, pages 595-596) of the genericity of a finite set of equilibria in a competitive exchange economy. It is also similar to the proof of Theorem 1 in Doraszelski and Escobar (2010) though our model is different because it is static and with incomplete information. The basic mathematical tools for our proof are the Transversality theorem, the Hotz-Miller inversion theorem, and the Implicit Function Theorem. We start enunciating the Transversality theorem.

Transversality theorem (Mas-Colell et al., 1995; Proposition 17.D.3, page 595). Let $\mathcal{U} \subseteq \mathbb{R}^n$ and $\Theta \subseteq \mathbb{R}^q$ be open sets, and let $f : \Theta \times \mathcal{U} \to \mathbb{R}^n$ be continuously differentiable. Suppose that the Jacobian matrix $\partial f(\theta, \mathbf{u})/\partial(\theta', \mathbf{u}')$ has rank n for all values $(\theta, \mathbf{u}) \in \Theta \times \mathcal{U}$ with $f(\theta, \mathbf{u}) = 0$. Then, for almost every $\theta \in \Theta$ the Jacobian $\partial f(\theta, \mathbf{u})/\partial \mathbf{u}'$ has rank n for every $\mathbf{u} \in \mathcal{U}$ with $f(\theta, \mathbf{u}) = 0$.

We now proceed with the proof of Lemma 1. Define $n \equiv NJ$ and $q = NJ(J+1)^{N-1}$, let $\mathcal{P} \subset \mathbb{R}^n$ be the open set of players' choice probabilities $\mathbf{P}$, and let $\Pi \subset \mathbb{R}^q$ be the open set of payoff parameters $\boldsymbol{\pi}$. Our equilibrium mapping $f(\boldsymbol{\pi}, \mathbf{P})$ is $f : \Pi \times \mathcal{P} \to \mathbb{R}^n$, and it is continuously differentiable.

Proof Lemma 1(A). It proceeds in four parts, [i] to [iv].
[i] Remember that $\Psi(\boldsymbol{\pi}, \mathbf{P}) = \{\Psi_i(\boldsymbol{\pi}_i, \mathbf{P}_{-i}) : i \in \times I\}$ with $\Psi_i(\boldsymbol{\pi}_i, \mathbf{P}_{-i}) = \{\Psi_i(a_i|\boldsymbol{\pi}_i, \mathbf{P}_{-i}) : a_i \neq 0\}$ and where $\Psi_i(a_i|\boldsymbol{\pi}_i, \mathbf{P}_{-i})$ is the choice probability for alternative a_i in the Random Utility Model (RUM) $a_i = \arg\max_{k \in \mathcal{A}} [u_i(k) + \varepsilon_i(k)]$ with $u_i(k) \equiv \sum_{\boldsymbol{a}_{-i}} (\prod_{j \neq i} P_j(a_j))\ \pi_i(k, \boldsymbol{a}_{-i})$. Note that $u_i(0) = 0$ because $\pi_i(0, \boldsymbol{a}_{-i}) = 0$. Let $\mathbf{u}_i$ be the $J \times 1$ vector of utilities $\{u_i(k) : k \neq 0\}$. We can represent the vector-valued best response function $\Psi_i(\boldsymbol{\pi}_i, \mathbf{P}_{-i})$ as a function from the vector of J utilities $\mathbf{u}_i$ into a vector of J choice probabilities $\mathbf{P}_i$. We can make this structure more explicit by representing function $\Psi_i(\boldsymbol{\pi}_i, \mathbf{P}_{-i})$ as $\Lambda_i(\mathbf{u}_i)$ where function $\Lambda_i(.)$ is the function that maps utility (differences) into choice probabilities in a RUM. The form of function Λ_i depends of the joint CDF of the random components ε_i. Hotz-Miller (1993, Proposition 1) establish that if the CDF of the vector ε_i is continuously differentiable with respect to the Lebesgue measure in the Euclidean space $\mathbb{R}^J$ (our Assumption 1B), then function $\Lambda_i(.)$ is invertible at every point $\mathbf{u}_i$. That is, for any player i and for any $\mathbf{u}_i \in \mathbb{R}^J$, the Jacobian matrix is $D_{\mathbf{u}_i}\Lambda_i$ is non-singular, i.e., its rank is equal to J.

[ii] Let $\boldsymbol{\pi}_i \equiv \{\pi_i(a_i, \boldsymbol{a}_{-i}) : a_i \neq 0,\ \boldsymbol{a}_{-i} \in \mathcal{A}^{N-1}\}$ be the vector of payoffs for player i. For this proof, it is convenient to represent this vector in terms of two subvectors: $\boldsymbol{\pi}_i = (\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)$ where $\boldsymbol{\alpha}_i$ is a vector of dimension J defined as $\boldsymbol{\alpha}_i = \{\pi_i(a_i, \mathbf{0}) : a_i \neq 0\}$, and $\boldsymbol{\beta}_i$ is a vector of dimension $J[(J+1)^{N-1} - 1]$ defined as $\boldsymbol{\beta}_i = \{\pi_i(a_i, \boldsymbol{a}_{-i}) - \pi_i(a_i, \mathbf{0}) : a_i \neq 0,\ \boldsymbol{a}_{-i} \in \mathcal{A}^{N-1}, \boldsymbol{a}_{-i} \neq \mathbf{0}\}$. That is, $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)$ is reparameterization of $\boldsymbol{\pi}_i$, and this reparameterization is without loss of generality. Using these definitions, taking into account the structure of the utilities $u_i(a_i)$ defined in point [i] above, and given that $\prod_{j \neq i} P_j(0) = 1 - \sum_{\boldsymbol{a}_{-i} \neq \mathbf{0}} (\prod_{j \neq i} P_j(a_j))$, we have that $u_i(a_i) = \alpha_i(a_i) + \sum_{\boldsymbol{a}_{-i} \neq \mathbf{0}} (\prod_{j \neq i} P_j(a_j))\ \beta_i(a_i, \boldsymbol{a}_{-i})$, or in vector form $\mathbf{u}_i = \boldsymbol{\alpha}_i + \mathbf{H}(\mathbf{P}_{-i})\ \boldsymbol{\beta}_i$, where $\mathbf{H}(\mathbf{P}_{-i})$ is a matrix with probabilities $\prod_{j \neq i} P_j(a_j)$. Using this notation, we have that $\Psi_i(\boldsymbol{\pi}_i, \mathbf{P}_{-i}) = \Lambda_i(\mathbf{u}_i) = \Lambda_i(\boldsymbol{\alpha}_i + \mathbf{H}(\mathbf{P}_{-i})\ \boldsymbol{\beta}_i)$. Given this structure, it is clear that the Jacobian matrix $D_{\boldsymbol{\alpha}_i}\Psi_i$ is

equal to the Jacobian $D_{\mathbf{u}_i}\Lambda_i$. Therefore, by the result [i] above, we have that for any player i and for any $(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, \mathbf{P}_{-i})$ the Jacobian matrix $D_{\boldsymbol{\alpha}_i}\Psi_i$ is non-singular and it has rank J.

[iii] Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be the vectors with the $\boldsymbol{\alpha}_i$'s and $\boldsymbol{\beta}_i$'s for every player such that $\boldsymbol{\pi} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$. We can represent $f(\boldsymbol{\pi}, \mathbf{P})$ as $f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P})$ with $f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P}) = \mathbf{P} - \{\Lambda_i(\boldsymbol{\alpha}_i + \mathbf{H}(\mathbf{P}_{-i})\,\boldsymbol{\beta}_i) : i \in \mathcal{I}\}$. Since the vector of payoffs $\boldsymbol{\alpha}_i$ enters in the best response function of player i, Λ_i, but not in the best responses of players other than i, we have that the Jacobian matrix $D_{\boldsymbol{\alpha}} f \equiv \partial f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P})/\partial \boldsymbol{\alpha}'$ is a block diagonal matrix where each block corresponds to $-D_{\boldsymbol{\alpha}_i}\Psi_i$ for each player i. Therefore, the previous result in [ii] that $rank[D_{\boldsymbol{\alpha}_i}\Psi_i] = J$ implies that at every point $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P})$ the rank of $D_{\boldsymbol{\alpha}} f$ is equal to $n = NJ$. In turn, this implies that the Jacobian matrix $D_{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{P}} f \equiv \partial f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P})/\partial(\boldsymbol{\alpha}', \boldsymbol{\beta}', \mathbf{P}')$ has rank $n = NJ$ for any value $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P})$.

[iv] Previous result [iii] provides the condition we need to apply the Transversality theorem to our model. Therefore, by the Transversality theorem, we have that for almost every $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ the Jacobian $\partial f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P})/\partial(\boldsymbol{\alpha}', \boldsymbol{\beta}', \mathbf{P}')$ has rank NJ for every $\mathbf{P} \in \mathcal{P}$ with $f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{P}) = 0$. That is, for almost every vector of payoffs $\boldsymbol{\pi}$ the set of equilibria $\Gamma(\boldsymbol{\pi})$ includes only regular equilibria.

Proof Lemma 1(B). Given $\boldsymbol{\pi}$, an equilibrium $\mathbf{P}$ is *locally unique* or *isolated* if there is a small scalar $\delta > 0$ such that if $\mathbf{P}' \neq \mathbf{P}$ and $\|\mathbf{P}' - \mathbf{P}\| < \delta$ then $f(\boldsymbol{\pi}, \mathbf{P}') \neq 0$ (Mas-Colell et al., 1995; page 590). First, we show that the Implicit Function Theorem implies that every regular equilibrium is isolated. For a regular equilibrium, the non-singularity of the Jacobian $D_{\mathbf{P}} f$ implies that $D_{\mathbf{P}} f$ $d\mathbf{P} \neq 0$ for any vector of infinitesimal changes $d\mathbf{P}$. Therefore, keeping $\boldsymbol{\pi}$ constant, we cannot remain in equilibrium if $\mathbf{P}' = \mathbf{P} + d\mathbf{P}$, i.e., given $\boldsymbol{\pi}$ fixed, a regular equilibrium $\mathbf{P}$ is isolated. This implies that if the set of equilibria $\Gamma(\boldsymbol{\pi})$ contains only regular equilibria, then this set is discrete. Now, given that the set of CCPs $\mathcal{P}$ is $(0,1)^{NJ}$ that is bounded, we have that the set of regular equilibria $\Gamma(\boldsymbol{\pi})$ is discrete and bounded, and a discrete and bounded set is finite (Theorem M.F.3 in section M.F. of the mathematical appendix in Mas-Colell et al., 1995).

Proof Lemma 1(C). Let $(\boldsymbol{\pi}^0, \mathbf{P}^0)$ be a regular equilibrium. The Implicit Function Theorem applied to $f(\boldsymbol{\pi}^0, \mathbf{P}^0) = 0$ implies that there is a continuously differentiable function that relates $\mathbf{P}$ and $\boldsymbol{\pi}$ in a neighborhood of the equilibrium $(\boldsymbol{\pi}^0, \mathbf{P}^0)$. Suppose that $(\boldsymbol{\pi}^1, \mathbf{P}^1)$ is an equilibrium that belongs to this neighborhood. Then, by continuity of the implicit function, there is a continuous path $\{\mathbf{P}[t] : t \in [0,1]\}$ with $f(\ [1-t]\ \boldsymbol{\pi}^0 + t\ \boldsymbol{\pi}^1,\ \mathbf{P}[t]) = 0$ for every $t \in [0,1]$. By the definition of equilibrium type, this means that $(\boldsymbol{\pi}^0, \mathbf{P}^0)$ and $(\boldsymbol{\pi}^1, \mathbf{P}^1)$ belong to the same type. The equilibrium type of $(\boldsymbol{\pi}^0, \mathbf{P}^0)$ is described by the set of equilibria in the neighborhood of $(\boldsymbol{\pi}^0, \mathbf{P}^0)$ for which the implicit function exits. ■

APPENDIX: 2 x 2 x 2 Game

This Appendix presents the expressions for the Hessian matrix $\nabla^2 \ell^0$ and the Jacobian matrices $\nabla_{\mathbf{P}'} \mathbf{c}^0$ and $\mathbf{A}^0$ for the 2 x 2 x 2 game. We can represent the log-likelihood function $\ell^x(\mathbf{x}; h_A(\mathbf{x}), \mathbf{P}(\mathbf{x}))$ as:

$$\ell^x(\mathbf{x}; h_A(\mathbf{x}), \mathbf{P}(\mathbf{x})) = \sum_{(a_1,a_2)\in\{0,1\}^2} Q(a_1, a_2|\mathbf{x}) \ \ln\left[Q^\theta(a_1, a_2 \mid \mathbf{x}; h_A(\mathbf{x}), \mathbf{P}(\mathbf{x}))\right] \tag{A.1}$$

where $Q^\theta(.)$ represents the distribution of (a_1, a_2) conditional to $\mathbf{x}$ predicted by the model,

$$Q^\theta(a_1, a_2|\mathbf{x}; h_A(\mathbf{x}), \mathbf{P}(\mathbf{x})) \equiv h_A(\mathbf{x}) \ d(a_1, P_{1A}(\mathbf{x})) \ d(a_2, P_{2A}(\mathbf{x})) + (1 - h_A(\mathbf{x})) \ d(a_1, P_{1B}(\mathbf{x})) \ d(a_2, P_{2B}(\mathbf{x})) \tag{A.2}$$

with $d(a, P) \equiv [P]^a [1 - P]^{1-a}$. For notational simplicity, in the expressions below we omit $\mathbf{x}$ as an argument, but we refer all the time to the log-likelihood $\ell^x(\mathbf{x}; h_A(\mathbf{x}), \mathbf{P}(\mathbf{x}))$.

In the likelihood $\ell^x(h_A, \mathbf{P})$, the vector of scores is:

$$\frac{\partial \ln Q^\theta(a_1, a_2 \mid h_A, \mathbf{P})}{\partial [h_A, \mathbf{P}]} = \frac{1}{Q^\theta(a_1, a_2|h_A, \mathbf{P})} \begin{bmatrix} d(a_1, P_{1A}) \ d(a_2, P_{2A}) - d(a_1, P_{1B}) \ d(a_2, P_{2B}) \\ h_A \ (2a_1 - 1) \ d(a_2, P_{2A}) \\ (1 - h_A) \ (2a_1 - 1) \ d(a_2, P_{2B}) \\ h_A \ d(a_1, P_{1A}) \ (2a_2 - 1) \\ (1 - h_A) \ d(a_1, P_{1B}) \ (2a_2 - 1) \end{bmatrix} \tag{A.3}$$

The Hessian, evaluated at true parameter values, is equal to negative expected value of outer product of the scores, that we can represent as:

$$-\nabla^2 \ell^{x,0} = \sum_{(a_1,a_2)\in\{0,1\}^2} \frac{1}{Q(a_1, a_2)} \frac{\partial Q^\theta(a_1, a_2 \mid h_A^0, \mathbf{P}^0)}{\partial [h_A, \mathbf{P}]} \frac{\partial Q^\theta(a_1, a_2 \mid h_A^0, \mathbf{P}^0)}{\partial [h_A, \mathbf{P}]'} \tag{A.4}$$

The first column of the Hessian matrix is $-\sum_{(a_1,a_2)\in\{0,1\}^2} Q(a_1, a_2)^{-1} \ \partial Q^\theta / \partial [h_A, \mathbf{P}] \ \partial Q^\theta / \partial h_A$.

Applying equation (A.4), we have that:

$$
\begin{aligned}
-\nabla^2 \ell^{x,0}_{[h,P]h} &= \frac{[(1-P_{1A})\ (1-P_{2A}) - (1-P_{1B})\ (1-P_{2B})]}{Q(0,0)} \begin{bmatrix} (1-P_{1A})\ (1-P_{2A}) - (1-P_{1B})\ (1-P_{2B}) \\ -h_A\ (1-P_{2A}) \\ -(1-h_A)\ (1-P_{2B}) \\ -h_A\ (1-P_{1A}) \\ -(1-h_A)\ (1-P_{1B}) \end{bmatrix} \\
&+ \frac{[(1-P_{1A})\ P_{2A} - (1-P_{1B})\ P_{2B}]}{Q(0,1)} \begin{bmatrix} (1-P_{1A})\ P_{2A} - (1-P_{1B})\ P_{2B} \\ -h_A\ P_{2A} \\ -(1-h_A)\ P_{2B} \\ h_A\ (1-P_{1A}) \\ (1-h_A)\ (1-P_{1B}) \end{bmatrix} \\
&+ \frac{[P_{1A}\ (1-P_{2A}) - P_{1B}\ (1-P_{2B})]}{Q(1,0)} \begin{bmatrix} P_{1A}\ (1-P_{2A}) - P_{1B}\ (1-P_{2B}) \\ h_A\ (1-P_{2A}) \\ (1-h_A)\ (1-P_{2B}) \\ -h_A\ P_{1A} \\ -(1-h_A)\ P_{1B} \end{bmatrix} \\
&+ \frac{[P_{1A}\ P_{2A} - P_{1B}\ P_{2B}]}{Q(1,1)} \begin{bmatrix} P_{1A}\ P_{2A} - P_{1B}\ P_{2B} \\ h_A\ P_{2A} \\ (1-h_A)\ P_{2B} \\ h_A\ P_{1A} \\ (1-h_A)\ P_{1B} \end{bmatrix}
\end{aligned}
\tag{A.5}
$$

The vector $\nabla^2 \ell^{x,0}_{[h,P]h}$ has full column rank (i.e., rank 1) if and only if it has at least one element different to zero. Consider the first element of this vector:

$$
\begin{aligned}
-\nabla^2 \ell^{x,0}_{hh} &= \frac{[(1-P_{1A})\ (1-P_{2A}) - (1-P_{1B})\ (1-P_{2B})]^2}{Q(0,0)} \\
&+ \frac{[(1-P_{1A})\ P_{2A} - (1-P_{1B})\ P_{2B}]^2}{Q(0,1)} \\
&+ \frac{[P_{1A}\ (1-P_{2A}) - P_{1B}\ (1-P_{2B})]^2}{Q(1,0)} \\
&+ \frac{[P_{1A}\ P_{2A} - P_{1B}\ P_{2B}]^2}{Q(1,1)}
\end{aligned}
\tag{A.6}
$$

It is the sum of four components that are greater or equal to zero. Therefore, $\nabla^2 \ell^{x,0}_{hh}$ is equal to zero only if its four components are zero. It is straightforward to show that these four restrictions hold only if $P_{1A} = P_{1B}$ and $P_{2A} = P_{2B}$. Also, taking into account (A.5), we can see that if $P_{1A} = P_{1B}$ and $P_{2A} = P_{2B}$ then $\nabla^2 \ell^{x,0}_{[h,P]h}$ is a vector of zeroes. Therefore, $-\nabla^2 \ell^{x,0}_{[h,P]h}$ is full column rank if and only if $P_{iA} \neq P_{iB}$ for at least one of the players.

References

[1] Aguirregabiria, V. and P. Mira (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1-53.

[2] Allman, E. S., Matias, C., and Rhodes, J. A. (2009): "Identifiability of parameters in latent structure models with many observed variables," *The Annals of Statistics*, 37, 3099-3132.

[3] Altonji, J. and L. Segal (1996): "Small-sample bias in GMM estimation of covariance structures," *Journal of Business & Economic Statistics*, 14(3), 353-366.

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

[5] Arcidiacono, P. and R. Miller (2011): "CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 79, 1823–1867.

[6] Bajari, P., L. Benkard and J. Levin (2007): "Estimating dynamic models of imperfect competition," *Econometrica*, 75, 1331-1370.

[7] Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010): "Estimating Static Models of Strategic Interactions," *Journal of Business & Economic Statistics*, 28(4), 469-482.

[8] Bajari, P., H. Hong, and D. Nekipelov (2013): "Game theory and econometrics: a survey of some recent research," in Advances in Economics and Econometrics: Tenth World Congress, D. Acemoglu, M. Arellano, and E. Dekel (eds.). Cambridge University Press.

[9] Bayer, P. and C. Timmins (2005): "On the Equilibrium Properties of Locational Sorting Models," *Journal of Urban Economics*, 57, 462-477.

[10] Bayer, P. and C. Timmins (2007): "Estimating Equilibrium Models of Sorting across Locations," *The Economic Journal*, 117, 353-374.

[11] Bergemann, D., and S. Morris (2013): "Robust predictions in games with incomplete information," *Econometrica*, 81, 1251-1308.

[12] Bergemann, D., and S. Morris (2016): "Bayes correlated equilibrium and the comparison of information structures in games," *Theoretical Economics*, 11, 487-522.

[13] Cooper, R. (1999): "Coordination Games." Cambridge University Press. Cambridge, UK.

[14] Cooper, R., and D. Corbae (2002): "Financial collapse: A lesson from the great depression," *Journal of Economic Theory*, 107(2), 159-190.

[15] Davis, D. and D. Weinstein (2002): "Bones, Bombs, and Breakpoints: The Geography of Economic Activity," *American Economic Review*, 92(5), 1269–1289.

[16] Davis, D. and D. Weinstein (2008): "A Search for Multiple Equilibria in Urban Industrial Structure," *Journal of Regional Science*, 48(1), 29-65.

[17] Delgado, M., and J. Hidalgo (2000): "Nonparametric Inference on Structural Breaks," *Journal of Econometrics*, 96, 113-144.

[18] De Paula, A., and X. Tang (2012): "Inference of Signs of Interaction Effects in Simultaneous Games With Incomplete Information," *Econometrica*, 80(1), 143-172.

[19] Doraszelski, U. and J. Escobar (2010): "A theory of regular Markov perfect equilibria in dynamic stochastic games: Genericity, stability, and purification," *Theoretical Economics*, 5, 369–402.

[20] Egan, M., A. Hortaçsu, and G. Matvos (2917): "Deposit competition and financial fragility: Evidence from the us banking sector," *American Economic Review*, 107(1), 169-216.

[21] Ellickson, P. and S. Misra (2008): "Supermarket Pricing Strategies," *Marketing Science*, 27(5), 811-828.

[22] Farmer, R. and J-T Guo (1995): "The Econometrics of Indeterminacy: An Applied Study," *Carnegie-Rochester Conference Series on Public Policy*, 43, 225-271.

[23] Grieco, P. (2014): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets," *RAND Journal of Economics*, 45(2), 303-340.

[24] Hall, P., and X. Zhou (2003): "Nonparametric Estimation of Component Distributions in a Multivariate Mixture," *Annals of Statistics*, 31(1), 201-224.

[25] Hall, P., A. Neeman, R. Pakyari, and R. Elmore (2005): "Nonparametric Inference in Multivariate Mixtures," *Biometrika* , 92, 667-678.

[26] Haller, H. and R. Lagunoff (2000): "Genericity and Markovian behavior in stochastic games." *Econometrica*, 68, 1231–1248.

[27] Hahn, J. and R. Moon (2010): "Panel Data Models with Finite Number of Multiple Equilibria," *Econometric Theory*, 26(03), 863-881.

[28] Horowitz, J. (1998): "Bootstrap Methods for Covariance Structures," *Journal of Human Resources*, 33(1), 39-61.

[29] Hotz, V. J., and R. A. Miller (1993): "Conditional choice probabilities and the estimation of dynamic models," *The Review of Economic Studies*, 60(3), 497-529.

[30] Kasahara, H. and K. Shimotsu (2014): "Nonparametric Identification and Estimation of the Number of Components in Multivariate Mixtures," *Journal of the Royal Statistical Society - Series B*, 76(1), 97-111.

[31] Kleibergen, F. and R. Paap (2006): "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, 133(1), 97-126.

[32] Krugman, P. (1991): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483-499.

[33] Liu, N., Q. Vuong, and H. Xu (2013): "Rationalization and Identification of Discrete Games with Correlated Types," manuscript. New York University.

[34] Magnolfi, L., and C. Roncoroni (2017): "Estimation of Discrete Games with Weak Assumptions on Information," manuscript. University of Wisconsin-Madison.

[35] Mas-Colell, A., M. Whinston, and J. Green (1995): "Microeconomic theory," Oxford University Press. New York.

[36] Moro, A. (2003): "The Effect of Statistical Discrimination on Black-White Wage Inequality: Estimating a Model with Multiple Equilibria," *International Economic Review*, 44 (2), 467-500.

[37] Müller, H. (1992): "Change-Points in Nonparametric Regression Analysis," *Annals of Statistics*, 20, 737-761.

[38] Otsu, T., M. Pesendorfer, and Y. Takahashi (2016): "Pooling Data across Markets in Dynamic Markov Games," *Quantitative Economics*, 7(2), 523-559.

[39] Pesendorfer, M. and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimators for Dynamic Games," *Review of Economic Studies*, 75, 901–928.

[40] Sweeting, A. (2009): "The Strategic Timing of Radio Commercials: An Empirical Analysis Using Multiple Equilibria", *RAND Journal of Economics*, 40(4), 710-742.

[41] Todd, P., and K. I. Wolpin (2012): "Estimating a Coordination Game within the Classroom," manuscript. Department of Economics. University of Pennsylvania.

[42] Wan, Y. and H. Xu (2014): "Semiparametric identification of binary decision games of incomplete information with correlated private signals," *Journal of Econometrics*, 182(2), 235–246.