Sharp test for equilibrium uniqueness in discrete games with a flexible information structure

Mathieu Marcoux*

December 18, 2018

Abstract

I propose a test for an assumption commonly maintained when estimating static discrete games of incomplete information, i.e. the assumption of equilibrium uniqueness in the data generating process. The test is appealing for several reasons. It allows for discrete common knowledge payoff-relevant unobserved heterogeneity (henceforth unobserved heterogeneity, for short). In that sense, it is more general than tests which attribute all correlation between players' decisions to multiple equilibria. Furthermore, the test does not require the estimation of payoffs to separate multiplicity of equilibria from unobservable heterogeneity. It is therefore useful in empirical applications leveraging multiple equilibria to identify the model's primitives when commonly-used exclusion restrictions are not available. Finally, it makes no parametric assumption on the payoff functions nor the distribution of players' private information. The main identifying assumption is the existence of an observable variable that plays the role of a proxy for the unobserved heterogeneity. The procedure boils down to testing the emptiness of the set of data generating processes that can rationalize the sample through a single equilibrium and a finite mixture over unobserved heterogeneity. Simulation evidence is provided to study the test's properties.

Keywords: Static discrete game, equilibrium uniqueness, incomplete information, common knowledge unobservable information, finite mixtures, partial identification.

*Département de sciences économiques and CIREQ, Université de Montréal. Address: C.P. 6128, succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada. Tel: 514-343-2399. Email: mathieu.marcoux@umontreal.ca. I thank Victor Aguirregabiria, Andrés Aradillas-López, Vincent Boucher, Martin Burda, Christian Gouriéroux, Paul Grieco, Marc Henry, Yingyao Hu, Yao Luo, Ismael Mourifié, Joris Pinkse, Eduardo Souza Rodrigues, Yuanyuan Wan and Ruli Xiao for comments and insightful discussions. I benefited from the comments of several seminar participants at Penn State and University of Toronto. Finally, I thank Christopher Ferrall for useful coding advices. All errors are mine. Financial support from Social Sciences and Humanities Research Council (SSHRC) and Ontario Graduate Scholarship (OGS) is gratefully acknowledged.

1 Introduction

Economic models of strategic interactions among agents often admit multiple equilibria. Multiplicity of equilibria *in the model* may be seen as an *economic* problem and some equilibrium refinement can be used to determine which equilibrium or which equilibria should be considered. In empirical games, multiplicity of equilibria *in the data generating process* is an *econometric* issue that must be taken into account when trying to recover the model's primitives from the data. In fact, identification arguments available in the literature differ according to the assumptions maintained on the number of equilibria realized in a given sample. Testing for equilibrium uniqueness is therefore desirable to guide applied researchers towards an appropriate estimation approach. Furthermore, assumptions on the number of equilibria realized in the data have different implications depending on the information structure of the game, i.e. whether the unobservables from the econometrician's point of view are private information or common knowledge among players. It follows that tests allowing for a relatively more flexible information structure may be preferable in practice.

This paper provides a test of equilibrium uniqueness in the data generating process under a flexible information structure. It allows for private information as well as discrete common knowledge payoff-relevant unobservables (henceforth unobserved heterogeneity, for short). The identification argument I propose is nonparametric: no parametric assumptions are needed for the payoff functions, nor the distribution of private information shocks; but the distribution of common knowledge unobserved heterogeneity and the equilibrium selection mechanism are assumed to have discrete supports. More precisely, the observable joint distribution of players' decisions can be written as a finite mixture and I use partial identification results from Henry, Kitamura, and Salanié (2014) to derive sharp bounds for the distributions defining this finite mixture. Equilibrium uniqueness in the data imposes further restrictions on these distributions. Testing the null hypothesis of equilibrium uniqueness in the data generating process simply amounts to testing whether the identified set constructed from all these restrictions is empty. The identification result relies on the existence of an observable variable that can be interpreted as a proxy for the common knowledge unobservable heterogeneity. It must (i) have sufficient variation; (ii) be correlated with these common knowledge unobservables; and (iii) provide only redundant information about players' decisions and the equilibrium selection if such unobservables were actually observed. The test is implemented through a simple two-stage approach suggested by Shi and Shum (2015) and simulation results suggest that it performs well in finite samples.

How one treats multiple equilibria typically depends on how one is willing to interpret the information unobservable to the econometrician. In games where one assumes that all unobservables are known to all players, i.e. games of complete information, set-identified estimators have been proposed to recover the set of model's primitives that can rationalize the data for any possible equilibrium selection mechanism (e.g. Tamer, 2003; Ciliberto and Tamer, 2009; Beresteanu, Molchanov, and Molinari, 2011; Galichon and Henry, 2011; etc.). In that sense, such estimation methods are robust to multiplicity of equilibria. In contrast, many estimation methods ask the econometrician to take a stance on whether or not there are multiple equilibria in the data when estimating games of incomplete information, i.e. games assuming that unobservables are players' private information. On one hand, many estimation methods assume that the data have been generated by a single equilibrium (e.g., Aguirregabiria and Mira, 2007; Bajari, Benkard, and Levin, 2007; Pakes, Ostrovsky, and Berry, 2007; Pesendorfer and Schmidt-Dengler, 2008; Bajari, Hong, Krainer, and Nekipelov, 2010; Aradillas-López, 2012; etc.). This assumption is often labeled the "single equilibrium in the data" or the "degenerate equilibrium selection mechanism" assumption. On the other hand, multiple equilibria realized in the data provide an extra source of variation that helps to identify the primitives of the model (e.g., Sweeting, 2009; De Paula and Tang, 2012; Aradillas-López and Gandhi, 2016). Multiple equilibria is therefore a valuable alternative to commonly-used player-specific exclusion restrictions when the latter are not available in practice. In those cases, one should therefore explicitly identify the different equilibria realized in the data instead of assuming them away.

According to the single equilibrium in the data assumption, every time the same players play the same game, the same equilibrium is realized. For example, consider a game of market entry between two players, firm A and firm B. Suppose that this game has the following two equilibria: either A is more likely to enter the market than B, or *vice versa*. Such equilibria may arise in markets that are typically too small to justify simultaneous entry. In an econometric study of the entry behaviour of A and B, one would typically observe firms' entry decisions in several markets. The single equilibrium in the data assumption states that if A is more likely to enter than B in one specific market, then it also has to be more likely to enter than B whenever the same game is realized in another market. This assumption is maintained even if B being more likely to enter than A is also sustainable in equilibrium.

Of course, the single equilibrium in the data assumption substantially simplifies the estimation by avoiding the need of solving for all equilibria of the model: the only relevant equilibrium is the one realized in the data, and can therefore be estimated. However, if the assumption is falsely maintained, the resulting estimates are associated with a mixture of equilibria, which is typically not an equilibrium in itself.

Some tests of equilibrium uniqueness in the data have been proposed in the literature.[1] Two different approaches can be distinguished. The first one, would include tests proposed by De Paula and Tang (2012), Hahn, Moon, and Snider (2017) and Xiao (2018). These tests treat correlation in players' decisions as evidence against the single equilibrium in the data assumption. In other words, they require players' decisions to be mutually independent after controlling for observable common knowledge information and the selected equilibrium. This requirement implies that unobservables are players' independent private information or, in other words, players do not know more about each other than the econometrician does. An alternative explanation of correlation in players' decisions would be that the unobservables interpreted as private information shocks are actually, at least partially, observed by competitors (e.g., Navarro

[1]I am focusing on static games. For dynamic games, Otsu, Pesendorfer, and Takahashi (2016) and Luo, Xiao, and Xiao (2018) have proposed some tests which will be further discussed below.

and Takahashi, 2012). Therefore, in such tests, common knowledge unobservable heterogeneity is ruled out by assumption and more importantly, may lead to the false rejection of the null hypothesis of a single equilibrium in the data.

The second approach, which includes the paper by Aguirregabiria and Mira (2018), has the advantage of allowing for common knowledge unobserved heterogeneity. A test of equilibrium uniqueness in the data generating process can be obtained as a by-product of their identification results. A detailed discussion of the differences between their approach and the test I propose is given below. At this point, it is worth mentionning the main difference: I do not require estimating the payoff functions using commonly-used player-specific exclusion restrictions to separate the problems of multiple equilibria from common knowledge unobservable heterogeneity. This distinction is relevant for an applied researcher who does not observe such exclusion restrictions, but instead hopes to use multiple equilibria to identify the model's primitives (as in Sweeting, 2009). Of course, this advantage is not free. The test proposed in the current paper requires observing a proxy for the common knowledge unobserved heterogeneity. In that sense, it trades exclusion restrictions for a proxy variable. Examples of suitable candidates of proxies are discussed in Section 4.

The rest of the paper is organized as follows. Related literature is summarized in Section 2, with a special attention being paid to some useful results on the nonparametric identification of finite mixtures. A static discrete game with simultaneous decisions is introduced in Section 3. The nonparametric identification results and the statistical test are respectively presented in Sections 4 and 5. Section 6 concludes. Proofs and further details are included in Appendices.

2 Related literature

As mentioned above, a few papers from the static game literature propose tests of equilibrium uniqueness in the data generating process. These tests are usually obtained as by-products of identification results. While these identification results provide great insights, some caveats about the corresponding tests for equilibrium uniqueness have already been pointed out in Section 1. In fact, many of the existing tests (e.g., De Paula and Tang, 2012; Hahn, Moon, and Snider, 2017; Xiao, 2018) require players' equilibrium-specific decisions to be independent given the observable information, hence ruling out common knowledge unobserved heterogeneity. In a static game of pure incomplete information this simply follows from the conditional independence of unobservable private shocks. In this setting, testing for equilibrium uniqueness boils down to testing whether players' decisions are conditionally independent.

The same conditional independence is also key to use recent nonparametric identification results from the literature on finite mixtures and measurement errors (e.g., Hall and Zhou, 2003; Hu, 2008; Kasahara and Shimotsu, 2009, 2014; Hu and Shum, 2012; Bonhomme, Jochmans, and Robin, 2016). [2] In games of pure incomplete information, one can use results from this literature to identify a lower bound on the number of equilibria occurring in the sample, which

[2]For a recent survey of the applications of measurement error models to empirical industrial organization and labour economics, see Hu (2017).

corresponds to the number of components in the finite mixture representation of the joint distribution of players' decisions. This is the approach proposed by Xiao (2018).

Unfortunately, as mentioned above, even if there is a single equilibrium in the data, conditional independence breaks down if players also take into account payoff-relevant information that is known to all of them, but unobservable to the econometrician. In such cases, tests assuming conditional independence cannot be applied.[3] The main issue is that if one finds the correlation between players' decisions to be non-zero or if one finds more than one components in the finite mixture representing choice probabilities, it could either be due to multiple equilibria in the data and/or common knowledge unobserved heterogeneity. In other words, such unobservables may lead to the false rejection of the single equilibrium in the data hypothesis.

Some progress has been made to allow for both private information and common knowledge unobserved heterogeneity in empirical games. In particular, Grieco (2014) proposes a framework that allows for multiple equilibria and a flexible information structure in a parametric setting. His results suggest that both private and common knowledge unobservable information may be relevant in empirical applications. More recently, Magnolfi and Roncoroni (2017) propose an estimation method applicable to games defined via an alternative equilibrium concept (Bayes Correlated Equilibrium developed by Bergemann and Morris, 2016) that allows for weak assumptions on the information structure of the game. In their empirical application, they also find that assumptions maintained on the information structure have an important impact on parameters estimates and counterfactual predictions. Such recent practical insight justifies the need to extend tests of equilibrium uniqueness beyond the pure incomplete information setting.

To the best of my knowledge, the only semi-parametric identification results that allow for multiple equilibria and common knowledge unobserved heterogeneity in static games are due to Aguirregabiria and Mira (2018). Their main proposition follows from a sequential identification argument which combines results from the literature about nonparametric identification of finite mixtures. In a first step, they identify the nonparametric distribution of a discrete random variable with finite support that summarizes the information of the common knowledge unobservable heterogeneity and the unobservable variable that indicates which equilibrium is realized. Using the property that the equilibrium selection variable is payoff-irrelevant, the distribution of the unobservable heterogeneity and the equilibrium selection can be separately identified. In their setting, the number of equilibria corresponds to the cardinality of the support of the unobservable variable that selects the equilibrium realized in the data.

The sequential approach in Aguirregabiria and Mira (2018)'s main identification result may be problematic if one is solely interested in testing equilibrium uniqueness. There are two limitations worth pointing out. The first one has already been mentioned above. In their setting, one must estimate payoff functions to separate common knowledge unobserved heterogeneity from the variable indexing which equilibrium is realized. Doing so typically requires player-specific exclusion restrictions, i.e. variables that only affect a given player's decision through

[3] Xiao (2018) briefly discusses the idea of using her identification result to test for common knowledge unobserved heterogeneity at the end of her Section 2.2. However, this procedure cannot be used to test for equilibrium uniqueness conditional on common knowledge unobserved heterogeneity.

its beliefs about its competitors' behaviour (e.g. Bajari, Hong, Krainer, and Nekipelov, 2010). However, one important motivation for testing equilibrium uniqueness is when such exclusion restrictions are not available and one would like to leverage mulitiplicity of equilibria as an alternative source of variation to identify payoffs. In such cases, one cannot apply Aguirregabiria and Mira (2018)'s identification results to test for equilibrium uniqueness.

A second limitation, which also applies to other tests based on finite mixtures, is that the finite mixture framework restricts the number of components identifiable from the data. This restriction may not be innocuous. The largest number of components that one can identify in the first step of Aguirregabiria and Mira (2018)'s sequential argument is given by the number of alternatives available in the players' choice set, raised to the power $\lfloor \frac{N-1}{2} \rfloor$, where N is the number of players and $\lfloor \cdot \rfloor$ is the floor function. As a result, no mixture would be identifiable in a game of market entry between two players. Aguirregabiria and Mira (2018) also propose non-sequential results which require the exclusion restrictions commonly used in empirical games to be sufficiently over-identifying. While their non-sequential approach has some advantages (e.g., it may allow for $N = 2$), it is still not applicable without variables satisfying these exclusion resctrictions.

Of course, similar restrictions on the identifiable number of components are likely to arise in other identification arguments based on a finite mixture. The approach that I propose is no exception, but the conditions imposed here are less restrictive. A considerable advantage of the current paper's procedure is that the number of equilibria is not restricted prior to the test. The restriction only affects the support of the common knowledge unobservable heterogeneity.

At this point, it is worth distinguishing between two different approaches that have been proposed in the literature about nonparametric identification of finite mixtures: (i) the conditional independence, and (ii) the exclusion restriction (which will be interpreted as a proxy variable) approaches. In both cases, the main objective is to identify the number of components, the conditional component distributions and the mixing probability weights.

In the conditional independence approach, the joint distribution of observed variables conditional on the latent mixing variable can be factored as the product of its marginals. A system of equations is constructed by considering different sub-vectors of the vector of mixed variables. For instance, in the context of a game with N players, one could consider the joint distributions of all subsets of players' decisions. Point identification is reached if one can construct enough equations to identify all the corresponding marginal conditional component distributions and mixing probability weights. The conditional independence approach has been used by, among others, Hall and Zhou (2003), Kasahara and Shimotsu (2009), Hu and Shum (2012) and Bonhomme, Jochmans, and Robin (2016). This conditional independence is also needed to identify the number of mixtures when using Kasahara and Shimotsu (2014)'s results.

Alternatively, the exclusion restriction approach assumes that there exists an observable variable with sufficient variation that affects the mixing probability weights, but not the conditional component distributions.[4] With this restriction, one can write the conditional component

[4]Notice that this exclusion restriction does not correspond to the one commonly used to identify payoff functions as in Bajari, Hong, Krainer, and Nekipelov (2010). This is partly the reason why "proxy" will often be

distributions and the mixing probability weights in terms of some set-identified parameters. To my knowledge, Henry, Kitamura, and Salanié (2014) are the first to propose this approach and they also provide an extensive discussion suggesting that the required restriction often arises naturally in applied work. An especially relevant feature of this alternative approach is that the joint distribution conditional on the mixing variable does not have to be factorable in the product of its marginals. This is key in the context of the current paper: because the observable conditional choice probabilities are potentially mixed over multiple equilibria and common knowledge unobserved heterogeneity, players' decisions may fail to be independent after controlling for only one of these two types of unobservables.

It should be emphasized that the main object of interest in the current paper is the number of equilibria *in the data generating process*, which form a subset of the equilibria *in the model*. In that sense, the objective is very different from other works, such as Kasy (2015), focusing on the number of equilibria *in the model*. In fact, when applying his inference method to a game of incomplete information, Kasy (2015) first estimates the model using a two-step approach. Such two-step estimation relies on the single equilibrium in the data assumption, which can be tested using the method I propose.

Finally, it is worth pointing out that some progress has recently been made in testing for equilibrium uniqueness with common knowledge unobserved heterogeneity in dynamic games. See for instance Otsu, Pesendorfer, and Takahashi (2016, Section 3.5) and Luo, Xiao, and Xiao (2018). A nice feature of the dynamic setting is that the econometrician typically observes each market for more than one periods, which gives some traction when trying to separate time-invariant unobserved heterogeneity from multiple equilibria. Unfortunately, this extra dimension, i.e. time, is usually not available in the static case. For this reason, the test proposed in the current paper is not directly comparable with tests for equilibrium uniqueness in dynamic games. However, it is interesting to note that, even with this extra dimension, tests applicable to dynamic games also focus on discrete unobserved heterogeneity.

3 A static discrete game with simultaneous decisions

While the model could be described in a more general way, I focus on a simple static discrete game with simultaneous decisions: a binary game between two players with two realizations of the common knowledge unobservable heterogeneity. This simple $2 \times 2 \times 2$ case helps in building the intuition behind the identification result. In particular, the simplification makes it fairly easy to represent the corresponding identified set graphically and to show some interesting results. I first describe the economic model, i.e. the game as it is played by the players. I then turn to its econometric counterpart observed by the econometrician.

preferred to "exclusion restriction" when referring to this approach in the current paper, even if the identification result is in fact based on an exclusion restriction.

3.1 Economic model

Consider a game where $N = 2$ players, indexed by $i \in \{1, 2\}$, simultaneously choose from a binary choice set. Players' decisions are stored in a vector of random variables $\boldsymbol{\mathcal{Y}} = [\mathcal{Y}_1, \mathcal{Y}_2]'$ with realizations $\mathbf{y} \in [y_1, y_2]' \in \{0, 1\}^2 \equiv \mathscr{Y}^2$.[5]

Players' decisions are contingent on some state variables, which are separated into two categories, depending on whether they are observed by all players. Let $\boldsymbol{\mathcal{S}} = [\boldsymbol{\mathcal{S}}_1', \boldsymbol{\mathcal{S}}_2']'$ with realizations $\mathbf{s} = [\mathbf{s}_1', \mathbf{s}_2']' \in \mathscr{S}^2$ be some information that is common knowledge to both players. Furthermore, let $\boldsymbol{\mathcal{E}} = [\mathcal{E}_1, \mathcal{E}_2]'$ with realizations $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2]' \in \mathbb{R}^2$ be some private information. Let $G_{\mathcal{E}_i}(\cdot)$ denote the cumulative density function of $\mathcal{E}_i$. Because player i's opponent does not observe ε_i, this is a game of incomplete information.

Let $\pi_i(\cdot) : \mathscr{Y}^2 \times \mathscr{S} \times \mathbb{R} \mapsto \mathbb{R}$ be player i's payoff function. While the payoff of player i choosing $y_i = 0$ is normalized to 0, the payoff when choosing $y_i = 1$ is denoted by $\pi_i(y_{-i}, \mathbf{s}, \varepsilon_i)$, where $-i$ denotes player i's opponent. The following assumption is maintained on the payoff functions and the distributions of $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{E}}$.

Assumption 1 (State variables and payoffs)**.** *(i)* $\boldsymbol{\mathcal{S}}$, $\mathcal{E}_1$ and $\mathcal{E}_2$ are mutually independent. *(ii)* $G_{\mathcal{E}_1}(\cdot)$, $G_{\mathcal{E}_2}(\cdot)$ are common knowledge to both players and they are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$. *(iii)* $\pi_1(\cdot)$ and $\pi_2(\cdot)$ are common knowledge to both players.

The timing of the decision process is as follows. First, $\mathbf{s}$ and $\boldsymbol{\varepsilon}$ are realized. Even if players do not observe their opponent's private information, they can still form beliefs about their competitor's decision under Assumption 1. Then, both players simultaneously decide, i.e. $\mathbf{y}$ is realized and commonly observed. To sum up, at the time of the simultaneous decisions, player i's information set is:

$$\mathcal{I}_i = \{\mathbf{s}, \varepsilon_i, \pi_1(\cdot), \pi_2(\cdot), G_{\mathcal{E}_1}(\cdot), G_{\mathcal{E}_2}(\cdot)\}. \tag{1}$$

Player i's strategy is a function that maps the information set, $\mathcal{I}_i$, to the choice set, i.e. $\sigma_i(\cdot) : \mathcal{I}_i \mapsto \mathscr{Y}$. For a given strategy, the conditional choice probability of player i choosing $y_i = 1$ at a given $\mathbf{s} \in \mathscr{S}$ is:

$$p_i(\mathbf{s}) = \int \mathbb{1}\{\sigma_i(\mathcal{I}_i) = 1\}\, dG_{\mathcal{E}_i}(\varepsilon_i) \tag{2}$$

which can be interpreted as the beliefs of player i's opponent regarding player i's decision, when player i behaves according to strategy $\sigma_i(\mathcal{I}_i)$. Collect those probabilities in $\mathbf{p}(\mathbf{s}) \equiv [p_1(\mathbf{s}), p_2(\mathbf{s})]'$. Using these choice probabilities, one can write the expected payoff of player i choosing $y_i = 1$ as:

$$\pi_i^{\mathbf{P}}(\mathbf{s}, \varepsilon_i) \equiv p_{-i}(\boldsymbol{s})\, \pi_i(1, \mathbf{s}, \varepsilon_i) + [1 - p_{-i}(\boldsymbol{s})]\, \pi_i(0, \mathbf{s}, \varepsilon_i). \tag{3}$$

[5]The following notation will be used in the current paper. Capital calligraphic letters are random variables. Their realizations are denoted by lower case roman letters. Script letters are used for the corresponding supports and $|\cdot|$ is the cardinality of this support. Boldface letters denote vectors and matrices.

If each player's strategy is to maximize expected payoffs, (2) can be written as:

$$p_i(\mathbf{s}) = \int \mathbb{1}\left\{\pi_i^{\mathbf{p}}(\mathbf{s}, \varepsilon_i) \geq 0\right\} dG_{\mathcal{E}_i}(\varepsilon_i). \quad (4)$$

The right hand side of equations (4) is the best response mapping of player i given its beliefs regarding its opponent's decision. Let $\psi_i(\cdot)$ denote this mapping and define $\mathbf{\Psi}(\cdot) \equiv [\psi_1(\cdot), \psi_2(\cdot)]'$. It follows that $\mathbf{p}(\mathbf{s})$ can be written as:

$$\mathbf{p}(\mathbf{s}) = \mathbf{\Psi}(\mathbf{s}, \mathbf{p}(\mathbf{s})) \quad (5)$$

where $\mathbf{\Psi}(\mathbf{s}, \cdot) : [0,1]^2 \mapsto [0,1]^2$ is a mapping in the probability space. Given this best response mapping, one can define a Bayesian Nash Equilibrium (BNE) in pure strategies. Defining a BNE in the probability space is very convenient to analyze equilibrium existence and multiplicity (Milgrom and Weber, 1985).

Definition 1 (BNE in probability space)**.** A pure strategy BNE in the probability space is a set of conditional choice probabilities $\mathbf{p}^*(\mathbf{s})$ such that $\mathbf{p}^*(\mathbf{s}) = \mathbf{\Psi}(\mathbf{s}, \mathbf{p}^*(\mathbf{s}))$.

Definition 1 simply states that, in equilibrium, players' beliefs are consistent with their opponent's. In fact, a BNE in the probability space is a fixed point of the best response mapping. Since $\mathbf{\Psi}(\mathbf{s}, \cdot)$ maps a compact set to itself and since it is continuous in $\mathbf{p}(\mathbf{s})$, the existence of an equilibrium follows from Brouwer's fixed point theorem for any $\mathbf{s} \in \mathbb{S}$. However, uniqueness is not guaranteed.

Let $\mathcal{T}$ be a random variable labelling which equilibrium is played. More precisely, each equilibrium is indexed by a realization $\tau \in \mathbb{T}(\mathbf{s})$ with conditional probability mass function $\lambda(\tau|\mathbf{s})$ which can be interpreted as the equilibrium selection mechanism given the information observable to both players.[6] The following assumption is maintained on $\mathcal{T}$.

Assumption 2 (Equilibrium index)**.** (i) $\mathcal{T}$ is independent from $\mathcal{E}_1$ and $\mathcal{E}_2$. (ii) $\mathcal{T}$ is discrete with support $\mathbb{T}(\mathbf{s}) \equiv \{1, \ldots, |\mathbb{T}(\mathbf{s})|\}$.

In order to fix ideas, I now introduce a running example that will be used as a data generating process to illustrate several concepts and results throughout the paper. This example is a simple static game of market entry between two firms.

Example 1 (Simple game of market entry)**.** Consider two firms deciding whether they want to operate in a given market, such that $y_i = 1$ if firm i enters the market and $y_i = 0$ otherwise. In this case, $\boldsymbol{\mathcal{S}}$ could be some common knowledge information about market's size and consumers' preferences. Moreover, $\boldsymbol{\mathcal{E}}$ could refer to some private information cost shifters such as managerial ability. Let players' payoffs when entering the market be:

$$\pi_1(y_2, \mathbf{s}, \varepsilon_1) = s_1 - 4y_2 - \varepsilon_1 \quad (6)$$

[6] A more rigorous notation would be to write $\mathcal{T}(\mathbf{s})$ and $\tau(\mathbf{s})$ instead of $\mathcal{T}$ and τ, but the latter is preferred to simplify notation.

$$\pi_2\left(y_1, \mathbf{s}, \varepsilon_2\right) = s_2 - 3y_1 - \varepsilon_2. \tag{7}$$

Furthermore, let $\boldsymbol{\mathcal{E}} \sim \text{Normal}\left(\mathbf{0}, \mathbf{I}_2\right)$, where $\mathbf{I}_2$ is the 2×2 identity matrix. Using $\Phi\left(\cdot\right)$ to denote the standard normal cumulative density function, the best response mapping is:

$$\mathbf{p}\left(\mathbf{s}\right) \equiv \begin{bmatrix} p_1\left(\mathbf{s}\right) \\ p_2\left(\mathbf{s}\right) \end{bmatrix} = \begin{bmatrix} \Phi\left(s_1 - 4p_2\left(\mathbf{s}\right)\right) \\ \Phi\left(s_2 - 3p_1\left(\mathbf{s}\right)\right) \end{bmatrix} \equiv \boldsymbol{\Psi}\left(\mathbf{s}, \mathbf{p}\left(\mathbf{s}\right)\right). \tag{8}$$

Figure 1, is the graphical representation of the best response mapping in (8). The BNE(s) are given by the intersection(s) of the two best response functions. This figure clearly illustrates that, for a given set of primitives, different realizations of $\boldsymbol{\mathcal{S}}$ are associated with different BNE's and in particular with different numbers of BNE's. ■

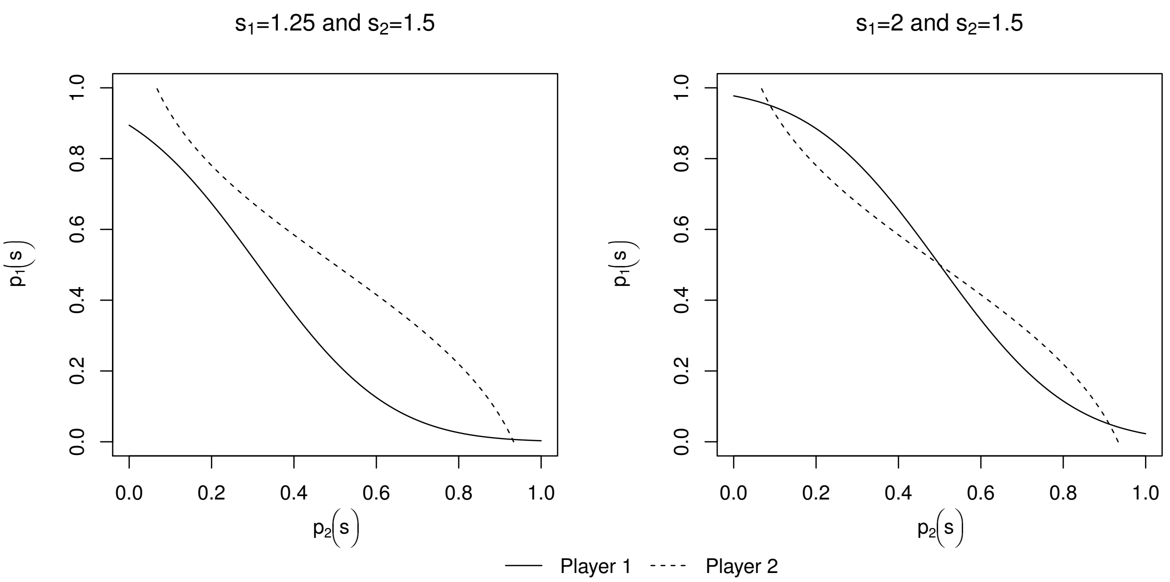

Figure 1: Equilibria multiplicity in Example 1

3.2 Econometric model

The game described so far is the *economic game* as it is played by the players. We now turn to the *econometric game*, i.e. the game as it is observed by the econometrician. An important difference between the two is that the researcher only observes some of the common knowledge payoff-relevant state variables in $\boldsymbol{\mathcal{S}}$. The following assumption is maintained on $\boldsymbol{\mathcal{S}}$.

Assumption 3 (Common knowledge payoff-relevant state variables)**.** (i) $\boldsymbol{\mathcal{S}} = \left[\boldsymbol{\mathcal{X}}', \boldsymbol{\mathcal{V}}'\right]'$, where $\boldsymbol{\mathcal{X}}$ is observable to the econometrician, but $\boldsymbol{\mathcal{V}}$ is not. (ii) $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{V}}$ have finite and discrete supports $\mathcal{X} = \left\{\mathbf{x}^0, \mathbf{x}^1, \ldots, \mathbf{x}^{|\mathcal{X}|-1}\right\}$ and $\mathcal{V}\left(\mathbf{x}\right) = \left\{\boldsymbol{\nu}^0, \boldsymbol{\nu}^1, \ldots, \boldsymbol{\nu}^{|\mathcal{V}(\mathbf{x})|-1}\right\}$. (iii) Realized $\boldsymbol{\nu}$'s are drawn from the conditional distribution with probability mass function $\Gamma\left(\cdot|\, \mathbf{x}\right)$. (iv) Conditional on $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{V}}$ is independent from $\boldsymbol{\mathcal{E}}$.

Typically, $\boldsymbol{\mathcal{V}}$ can be thought of as a vector of common knowledge unobservable heterogeneity.[7] In particular, $\boldsymbol{\mathcal{V}}$ could be a vector of player-specific unobservables. In the simple $2 \times 2 \times 2$ case, $|\mathcal{V}(\mathbf{x})| = 2$ and $\mathcal{V}(\mathbf{x}) = \{\boldsymbol{\nu}^0, \boldsymbol{\nu}^1\}$.

While $\mathcal{X}$ being a finite and discrete set is not essential in theory, it is very convenient in practice since the identification result holds for fixed values of $\mathbf{x}$. The applied researcher would find it helpful to discretize any continuous variables in $\boldsymbol{\mathcal{X}}$. Assumption 3 allows for a fairly flexible information structure. Notice that $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{V}}$ are not assumed to be independent across markets. This will be useful when justifying the choice of proxy variables proposed below.

Example 1 (Simple game of market entry, continued)**.** In the simple game of market entry between two firms, $\boldsymbol{\mathcal{V}}$ could refer to consumers' preferences in a given market. While both firms may have gathered information about these preferences through market research, such information is typically unobservable to the econometrician. Since consumers' preferences may affect firms' payoffs differently, the vector $\boldsymbol{\mathcal{V}}$ is allowed to have firm-specific components such that $\boldsymbol{\mathcal{V}} = [\mathcal{V}_1, \mathcal{V}_2]'$. The discrete support of $\boldsymbol{\mathcal{V}}$ can account for one firm being preferred to the other. For instance, one could have $\mathcal{V}(\mathbf{x}) = \{\boldsymbol{\nu}^0, \boldsymbol{\nu}^1\}$ such that $\nu_1^0 > \nu_2^0$ and $\nu_1^1 < \nu_2^1$. Moreover, $\boldsymbol{\mathcal{V}}$ varying with $\mathbf{x}$ captures that consumer's preferences may vary with observed market characteristics. ■

Another important difference between the game played by the players and the game observed by the econometrician is that not all equilibria in the model need to be realized in the data. Let $\mathcal{T}^*(\mathbf{x}, \boldsymbol{\nu}) \subseteq \mathcal{T}(\mathbf{x}, \boldsymbol{\nu})$ be the subset of the model's equilibria that are realized in the data according to probabilities given by $\lambda^*(\cdot|\, \mathbf{x}, \boldsymbol{\nu})$. Notice that $1 \leq |\mathcal{T}^*(\mathbf{x}, \boldsymbol{\nu})| \leq |\mathcal{T}(\mathbf{x}, \boldsymbol{\nu})|$ and the special case $|\mathcal{T}^*(\mathbf{x}, \boldsymbol{\nu})| = 1$ for all $\boldsymbol{\nu} \in \mathcal{V}(\mathbf{x})$ corresponds to the single equilibrium in the data assumption, conditional on the observable $\mathbf{x}$.

When estimating an empirical game, the econometrician typically observes M independent realizations of the game. For each of these realizations, the data consist in:

$$\{\mathbf{y}_m, \mathbf{x}_m : m = 1, \ldots, M\}. \tag{9}$$

There are therefore three random variables that are unobservable from the point of view of the researcher: (i) the private information shocks $\boldsymbol{\mathcal{E}}$; (ii) the common knowledge variables $\boldsymbol{\mathcal{V}}$; and (iii) the variable indicating which equilibrium is realized in the data $\mathcal{T}$.

Let $p(\mathbf{y}|\mathbf{x})$ be the observable conditional joint distribution of players' decisions. Given Assumptions 1 to 3, such distributions are double finite mixtures of the equilibrium conditional choice probabilities realized in the data, denoted $p^*(\mathbf{y}|\mathbf{x}, \boldsymbol{\nu}, \tau)$. For a given $(\mathbf{y}, \mathbf{x}) \in \mathcal{Y}^2 \times \mathcal{X}$:

$$p(\mathbf{y}|\mathbf{x}) = \sum_{\boldsymbol{\nu} \in \mathcal{V}(\mathbf{x})} p(\mathbf{y}|\mathbf{x}, \boldsymbol{\nu})\, \Gamma(\boldsymbol{\nu}|\mathbf{x}) \tag{10}$$

where

[7] Since the support of $\boldsymbol{\mathcal{V}}$ depends on $\mathbf{x}$, it may be preferable to write $\boldsymbol{\mathcal{V}}(\mathbf{x})$ and $\boldsymbol{\nu}(\mathbf{x})$. However, similarly as for $\mathcal{T}$ and τ, the argument is dropped to alleviate notation.

$$p(\mathbf{y}|\mathbf{x},\boldsymbol{\nu}) = \sum_{\tau \in \mathcal{T}^*(\mathbf{x},\boldsymbol{\nu})} p^*(\mathbf{y}|\mathbf{x},\boldsymbol{\nu},\tau)\,\lambda^*(\tau|\mathbf{x},\boldsymbol{\nu}). \tag{11}$$

From equations (10) and (11), one can see that $p(\mathbf{y}|\mathbf{x})$ consists in a double finite mixture with a total of $\sum_{j=0}^{|\mathcal{V}(\mathbf{x})|-1} |\mathcal{T}^*(\mathbf{x},\boldsymbol{\nu}^j)|$ components. In the current setting, the main object of interest is the number of equilibria associated with each $\boldsymbol{\nu} \in \mathcal{V}(\mathbf{x})$ conditional on $\mathbf{x}$. In the simple $2 \times 2 \times 2$ case, this object of interest corresponds to each element of the following set:

$$\left\{\left|\mathcal{T}^*\left(\mathbf{x},\boldsymbol{\nu}^j\right)\right|\right\}_{j=0}^{|\mathcal{V}(\mathbf{x})|-1} \equiv \left\{\left|\mathcal{T}^*\left(\mathbf{x},\boldsymbol{\nu}^0\right)\right|, \left|\mathcal{T}^*\left(\mathbf{x},\boldsymbol{\nu}^1\right)\right|\right\}. \tag{12}$$

4 Nonparametric identification

4.1 The need for more structure

Notice that, conditional on $\mathbf{x}$, $\boldsymbol{\nu}$ and τ, players' decisions are independent. This is a consequence of Assumption 1(i). Under some conditions on $|\mathcal{Y}^2|$, one can use recent results in the literature about the nonparametric identification of finite mixtures (Kasahara and Shimotsu, 2014), to identify the *total* number of components in the *double* finite mixture in (10) and (11), i.e. $\sum_{j=0}^{|\mathcal{V}(\mathbf{x})|-1} |\mathcal{T}^*(\mathbf{x},\boldsymbol{\nu}^j)|$. However, in the current setting, we want to test whether or not $|\mathcal{T}^*(\mathbf{x},\boldsymbol{\nu})| = 1\ \forall \boldsymbol{\nu} \in \mathcal{V}(\mathbf{x})$. We are therefore not necessarily interested in this total number of components.

Unfortunately, any combinations of $|\mathcal{V}(\mathbf{x})|$ and $\{|\mathcal{T}^*(\mathbf{x},\boldsymbol{\nu}^j)|\}_{j=0}^{|\mathcal{V}(\mathbf{x})|-1}$ that generate the same total number of components in the double finite mixture are observationally equivalent. Therefore, in order to separate $|\mathcal{V}(\mathbf{x})|$ from $\{|\mathcal{T}^*(\mathbf{x},\boldsymbol{\nu}^j)|\}_{j=0}^{|\mathcal{V}(\mathbf{x})|-1}$, one needs to put more structure on the double finite mixture of interest.

The additional structure that I impose is based on identification results from Henry, Kitamura, and Salanié (2014). While details about the argument are presented below, one can interpret the main requirement of this approach, when applied to the current setting, as the availability of a proxy variable for the common knowledge unobservable heterogeneity. Such proxy variable allows us to separate the mixture over $\boldsymbol{\nu}$ from the mixture over τ. More importantly, it does so without needing to estimate payoff functions.

However, one important drawback of their approach is that it leads to a partial identification result, which may be seen as unfortunate since other results currently available in the literature about nonparametric identification of finite mixtures deliver point identification. The reason why point identification fails in the current setting is that the finite mixture results from Henry, Kitamura, and Salanié (2014) are only used to deal with the mixure over common knowledge unobserved heterogeneity. However, conditional on $\mathbf{x}$ and $\boldsymbol{\nu}$, players' decisions are not independent if there are multiple equilibria realized in the data. Therefore, when identifying the finite mixture over $\boldsymbol{\nu}$, one cannot use the point identification results typically associated with the conditional independence approach.

In order to recover point identification, one could interpret the double finite mixture as a single finite mixture over a discrete variable that summarizes the information contained in both $\boldsymbol{\mathcal{V}}$ and $\mathcal{T}$. This approach is used by Aguirregabiria and Mira (2018) (as well as Luo, Xiao, and Xiao, 2018, in the dynamic case). Unfortunately, as already mentioned, such a finite mixture representation would put restrictions on the total number of elements in the support of the discrete variable. In other words, it potentially restricts $\sum_{j=0}^{|\mathcal{V}(\mathbf{x})|-1} |\mathcal{T}^* (\mathbf{x}, \boldsymbol{\nu}^j)|$. As a result, a test for equilibrium uniqueness based on such an identification approach would have the important drawback of restricting the number of equilibria that are identifiable prior to testing. By treating the identification of the mixture over $\boldsymbol{\nu}$ separately from the mixture over τ, the partial identification approach proposed here does not suffer from this limitation.

4.2 Identifying restrictions and proxy variable

The additional structure needed for identifying the mixture over the common knowledge unobserved heterogeneity is now stated in terms of exclusion restrictions, i.e. similarly as in Henry, Kitamura, and Salanié (2014). The proxy variable interpretation will be highlighted later on. Once again, it is worth emphasizing that the following exclusion restrictions are different from the ones commonly used for the identification of discrete games.

Let the vector of observable state variables, $\boldsymbol{\mathcal{X}}$, be divided into a sub-vector of variables that do not satisfy the exclusion restriction, $\boldsymbol{\mathcal{X}}_{\mathrm{NE}}$, and a subvector of variables that do satisfy it, $\boldsymbol{\mathcal{X}}_{\mathrm{E}}$, such that $\boldsymbol{\mathcal{X}} = [\boldsymbol{\mathcal{X}}'_{\mathrm{NE}}, \boldsymbol{\mathcal{X}}'_{\mathrm{E}}]'$, with realizations $\mathbf{x} = [\mathbf{x}'_{\mathrm{NE}}, \mathbf{x}'_{\mathrm{E}}]' \in \mathcal{X}_{\mathrm{NE}} \times \mathcal{X}_{\mathrm{E}} \equiv \left\{\mathbf{x}^0_{\mathrm{NE}}, \mathbf{x}^1_{\mathrm{NE}}, \ldots, \mathbf{x}^{|\mathcal{X}_{\mathrm{NE}}|-1}_{\mathrm{NE}}\right\} \times \left\{\mathbf{x}^0_{\mathrm{E}}, \mathbf{x}^1_{\mathrm{E}}, \ldots, \mathbf{x}^{|\mathcal{X}_{\mathrm{E}}|-1}_{\mathrm{E}}\right\}$. Assumption 4 formally states the identifying restrictions.

Assumption 4 (Identifying restrictions)**.** For any $\mathbf{y} \in \mathcal{Y}^2$, $\mathbf{x}_{\mathrm{NE}} \in \mathcal{X}_{\mathrm{NE}}$, $\mathbf{x}_{\mathrm{E}} \in \mathcal{X}_{\mathrm{E}}$ and $\boldsymbol{\nu} \in \mathcal{V}(\mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}})$: *(i)* (Support independence) $\mathcal{V}(\mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}})$ does not depend on $\mathbf{x}_{\mathrm{E}}$; *(ii)* (Cardinality of the support) $|\mathcal{V}(\mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}})| \leq \min\{|\mathcal{Y}|^2, |\mathcal{X}_{\mathrm{E}}|\}$; *(iii)* (Relevance) $\Gamma(\boldsymbol{\nu}|\mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}})$ depends on $\mathbf{x}_{\mathrm{E}}$; and *(iv)* (Redundancy) $p(\mathbf{y}|\mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}}, \boldsymbol{\nu})$ does not depend on $\mathbf{x}_{\mathrm{E}}$.

The support independence condition implies that the set of values of the mixing variables realized with a positive probability does not vary with $\mathbf{x}_{\mathrm{E}}$. This condition is important: in order to use variation in $\mathbf{x}_{\mathrm{E}}$ to identify the finite mixture over $\boldsymbol{\nu}$, such variation should not generate changes in the set of possible realizations of the mixing variable. Slightly abusing notation, $\mathcal{V}(\mathbf{x}_{\mathrm{NE}})$ will be used for the rest of the paper to make this condition explicit.

The condition on the cardinality of the support is included to make sure that there is enough variation in $\boldsymbol{\mathcal{Y}}$ and in $\boldsymbol{\mathcal{X}}_{\mathrm{E}}$ for the exclusion restriction approach to be able to capture all the relevant realizations of $\boldsymbol{\mathcal{V}}$. The rationale for $\min\{|\mathcal{Y}|^2, |\mathcal{X}_{\mathrm{E}}|\}$ will become clear in Lemma 1.

The relevance condition requires the distribution of the unobservable $\boldsymbol{\mathcal{V}}$ to depend on both $\boldsymbol{\mathcal{X}}_{\mathrm{NE}}$ and $\boldsymbol{\mathcal{X}}_{\mathrm{E}}$. More precisely, one needs different values of $\mathbf{x}_{\mathrm{E}}$ to be associated with different weights corresponding to each possible realization of $\boldsymbol{\mathcal{V}}$. Notice that this condition does not contradict the support independence one as long as realizations of $\boldsymbol{\mathcal{V}}$ do not become zero probability events for some values of $\mathbf{x}_{\mathrm{E}}$.

Finally, by the redundancy condition, the conditional choice probabilities, the equilibrium selection mechanism and the set of equilibria in the data generating process have to be independent of $\boldsymbol{\mathcal{X}}_{\mathrm{E}}$ after conditioning on $\boldsymbol{\mathcal{V}}$.[8] Once again, slightly abusing notation, such independence is made obvious by using $p\left(\mathbf{y}|\mathbf{x}_{\mathrm{NE}},\boldsymbol{\nu}\right)$, $p^*\left(\mathbf{y}|\mathbf{x}_{\mathrm{NE}},\boldsymbol{\nu},\tau\right)$, $\lambda^*\left(\tau|\mathbf{x}_{\mathrm{NE}},\boldsymbol{\nu}\right)$ and $\mathcal{T}^*\left(\mathbf{x}_{\mathrm{NE}},\boldsymbol{\nu}\right)$. In other words, $\boldsymbol{\mathcal{X}}_{\mathrm{E}}$ provides some information about the distribution of the unobservable $\boldsymbol{\mathcal{V}}$, but would not provide any information about players' decisions nor the equilibrium selection if $\boldsymbol{\mathcal{V}}$ would be observable.

Of course, a natural question to ask at this point is whether variables satisfying restrictions stated in Assumption 4 are easy to find. Remember that $\boldsymbol{\mathcal{V}}$ is observed by both players, but not by the econometrician. In some sense, one simply needs an *observable* variable that plays the role of a proxy for the *unobservable* common knowledge payoff-relevant variables. A suggestion of variables that can be used in the context of the simple market entry example is presented in Example 1 below.

Example 1 (Simple game of market entry, continued)**.** Let firm's payoffs be:

$$\pi_1\left(y_2,\mathbf{x}_{\mathrm{NE}},\nu_1,\varepsilon_1\right)=x_{\mathrm{NE},1}+\nu_1-4y_2-\varepsilon_1 \tag{13}$$

$$\pi_2\left(y_1,\mathbf{x}_{\mathrm{NE}},\nu_2,\varepsilon_2\right)=x_{\mathrm{NE},2}+\nu_2-3y_1-\varepsilon_2. \tag{14}$$

Provided that there is *some* level of correlation between the realizations of $\boldsymbol{\mathcal{V}}$ across markets, the realizations of $\boldsymbol{\mathcal{X}}_{\mathrm{NE}}$ in other markets can be used as a proxy for $\boldsymbol{\mathcal{V}}$. In other words, one can use some of the observable common knowledge state variables in surrounding markets to control for common knowledge unobserved heterogeneity in a given market. Here is the argument. Consider a market m and let m' denote some surrounding market(s). In each market, $\boldsymbol{\mathcal{V}}$ is correlated with $\boldsymbol{\mathcal{X}}_{\mathrm{NE}}$. For instance, the unobservable information that both firms may have regarding consumers' preferences in a given market is typically correlated with this market's observable demographics. Provided that $\boldsymbol{\nu}_{m'}$ is correlated with $\boldsymbol{\nu}_m$, then $\mathbf{x}_{\mathrm{NE},m'}$ contains some information about $\boldsymbol{\nu}_m$ through its correlation with $\boldsymbol{\nu}_{m'}$. $\mathbf{x}_{\mathrm{NE},m'}$ therefore satisfies the relevance condition. Furthermore, by definition of the BNE, only market m's state variables affect firms' decisions and equilibrium selection in this market. Therefore, $\mathbf{x}_{\mathrm{NE},m'}$ is only informative about market m when one fails to observe $\boldsymbol{\nu}_m$. In other words, $\mathbf{x}_{\mathrm{NE},m'}$ satisfies the redundancy requirement. Finally, a necessary condition for the support constraint to hold is that the support of $\mathbf{x}_{\mathrm{NE},m'}$ must be large enough to capture the different possible realizations $\boldsymbol{\nu}_m$, which can be checked for a given $|\mathcal{V}\left(\mathbf{x}_{\mathrm{NE}}\right)|$. ■

There is one potential caveat that should be pointed out. Even if being able to treat the mixture over $\boldsymbol{\nu}$ separately from the mixture over τ avoids restricting the number of equilibria

[8] Notice that the redundancy condition implies independence between $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{X}}_{\mathrm{E}}$ after conditioning on $\boldsymbol{\mathcal{X}}_{\mathrm{NE}}$ and $\boldsymbol{\mathcal{V}}$. In that sense, one may find the exclusion restriction approach for the identification of finite mixtures to be somewhat similar to the alternative approach relying on the independence of some observable variables after conditioning on the latent variable. However, if one was to leverage such independence between $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{X}}_{\mathrm{E}}$ to identify the mixture over $\boldsymbol{\mathcal{V}}$ as in the conditional independence approach, one would not be allowed to condition the distribution of the latent variable $\boldsymbol{\mathcal{V}}$ on $\boldsymbol{\mathcal{X}}_{\mathrm{E}}$.

in the data generating process, the exclusion restriction approach still puts an upper bound on $|\mathcal{V}(\mathbf{x}_{\text{NE}})|$ identifiable from the data.[9] The cardinality of the support condition in Assumption 4 assumes away the potential issue caused by the restriction on $|\mathcal{V}(\mathbf{x}_{\text{NE}})|$ implied by the finite mixture framework. Of course, one could simply impose a similar restriction in order to use Aguirregabiria and Mira (2018)'s identification results to test for equilibrium uniqueness in the data generating process. However, as mentioned above, the needed assumption would restrict both $|\mathcal{V}(\mathbf{x})|$ and $|\mathcal{T}^*(\mathbf{x}, \boldsymbol{\nu})|$ $\forall \boldsymbol{\nu} \in \mathcal{V}(\mathbf{x})$, instead of only $|\mathcal{V}(\mathbf{x})|$.

Even though we are now considering the simple case of $|\mathcal{V}(\mathbf{x}_{\text{NE}})| = 2$, it is worth noting that $|\mathcal{V}(\mathbf{x}_{\text{NE}})|$ is actually identifiable. Let $\mathbf{\Delta P}(\mathbf{x}_{\text{NE}})$ be the $(|\mathcal{X}_{\text{E}}| - 1) \times |\mathcal{Y}^2|$ matrix with element (k, j) given by $p\left(\mathbf{y}^j | \mathbf{x}_{\text{NE}}, \mathbf{x}_{\text{E}}^k\right) - p\left(\mathbf{y}^j | \mathbf{x}_{\text{NE}}, \mathbf{x}_{\text{E}}^0\right)$, for $k = 1, \ldots, |\mathcal{X}_{\text{E}}| - 1$ and $j = 1, \ldots, |\mathcal{Y}^2|$. As it is stated in Lemma 1, for each $\mathbf{x}_{\text{NE}} \in \mathcal{X}_{\text{NE}}$, $|\mathcal{V}(\mathbf{x}_{\text{NE}})|$ is identified through the rank of the matrix $\mathbf{\Delta P}(\mathbf{x}_{\text{NE}})$.

Lemma 1 (Identification of $|\mathcal{V}(\mathbf{x}_{\text{NE}})|$)**.** *Under Assumptions 1 to 4, for each* $\mathbf{x}_{\text{NE}} \in \mathcal{X}_{\text{NE}}$*,* $|\mathcal{V}(\mathbf{x}_{\text{NE}})| = \text{rank}\left\{\mathbf{\Delta P}(\mathbf{x}_{\text{NE}})\right\} + 1$.

Proof. This result follows from Henry, Kitamura, and Salanié (2014, Lemma 2, p. 138). Notice that they only consider mixtures of marginal distributions. However, their result also applies to mixtures of joint distributions, provided that one is interested in the joint conditional component distributions themselves, not the corresponding marginal distributions. □

Since the elements in matrix $\mathbf{\Delta P}(\mathbf{x}_{\text{NE}})$ are probabilities, its column rank is at most $|\mathcal{Y}^2| - 1$. Therefore, since the rank of a matrix is bounded by the minimum number of its rows and columns, the finite mixture representation restricts $|\mathcal{V}(\mathbf{x}_{\text{NE}})|$ to be at most $\min\{|\mathcal{Y}^2|, |\mathcal{X}_{\text{E}}|\}$, which corresponds to the cardinality of the support condition stated in Assumption 4.

4.3 Constructing the identified set

The identification result presented in this paper is conditional on $\mathbf{x}_{\text{NE}}$, such that all functions and statistics presented below depend on $\mathbf{x}_{\text{NE}}$. In order to alleviate notation, I omit $\mathbf{x}_{\text{NE}}$ as an argument.

Since we are considering the simple $2 \times 2 \times 2$ case, let $\Gamma(\mathbf{x}_{\text{E}}) \equiv \Gamma(\boldsymbol{\nu}^1 | \mathbf{x}_{\text{E}})$ such that $1 - \Gamma(\mathbf{x}_{\text{E}}) = \Gamma(\boldsymbol{\nu}^0 | \mathbf{x}_{\text{E}})$. Under Assumptions 1 to 4, (10) and (11) become:

$$p(\mathbf{y}|\mathbf{x}_{\text{E}}) = p\left(\mathbf{y}|\boldsymbol{\nu}^0\right)\left[1 - \Gamma(\mathbf{x}_{\text{E}})\right] + p\left(\mathbf{y}|\boldsymbol{\nu}^1\right)\Gamma(\mathbf{x}_{\text{E}}) \tag{15}$$

where

$$p\left(\mathbf{y}|\boldsymbol{\nu}^j\right) = \sum_{\tau \in \mathcal{T}^*(\boldsymbol{\nu}^j)} p^*\left(\mathbf{y}|\boldsymbol{\nu}^j, \tau\right)\lambda^*\left(\tau|\boldsymbol{\nu}^j\right). \tag{16}$$

Let $\boldsymbol{\theta} \equiv [\phi, \Upsilon]'$ where $\phi \equiv \Gamma(\mathbf{x}_{\text{E}}^0)$ and $\Upsilon \equiv \Gamma(\mathbf{x}_{\text{E}}^1) - \Gamma(\mathbf{x}_{\text{E}}^0)$. The main intuition behind the identification argument is as follows. Given the identifying restrictions stated above, one can

[9] More details on the number of components that can be identified are given below.

construct the identified set Θ_{I} such that each $\boldsymbol{\theta} \in \Theta_{\mathrm{I}}$ can rationalize the data through a finite mixture over $\boldsymbol{\nu} \in \mathcal{V}$ and a single equilibrium being realized for each $\boldsymbol{\nu} \in \mathcal{V}$, i.e. $|\mathcal{T}^*(\boldsymbol{\nu})| = 1$ $\forall \boldsymbol{\nu} \in \mathcal{V}$. For a given $\mathbf{x}_{\mathrm{NE}}$, testing $\Theta_{\mathrm{I}} \neq \emptyset$ boils down to checking whether testable implications of the single equilibrium in the data assumption hold.

To construct Θ_{I}, I closely follow the partial identification approach proposed by Henry, Kitamura, and Salanié (2014). More precisely, I rewrite the unknown probabilities $p(\mathbf{y}|\boldsymbol{\nu})$ and $\Gamma(\mathbf{x}_{\mathrm{E}})$ as functions of observable probabilities $p(\mathbf{y}|\mathbf{x}_{\mathrm{E}})$'s and $\boldsymbol{\theta}$. There are two types of restrictions that must be satisfied by $\boldsymbol{\theta}$ to belong to the identified set. In particular, $\boldsymbol{\theta} \in \Theta_{\mathrm{I}}$ if:

1. the resulting $p(\mathbf{y}|\boldsymbol{\nu})$'s and $\Gamma(\mathbf{x}_{\mathrm{E}})$'s belong to the unit interval; and

2. $p(\mathbf{y}|\boldsymbol{\nu})$'s satisfy the conditions implied by a single equilibrium being realized in the data.

The first type of restrictions are satisfied if the data can be rationalized through a finite mixture over common knowledge unobserved heterogeneity. In that sense, they correspond to the restrictions provided in Henry, Kitamura, and Salanié (2014). The second ones are extra restrictions that are satisfied if the single equilibrium in the data assumption holds for each $\boldsymbol{\nu} \in \mathcal{V}$ at a given $\mathbf{x}_{\mathrm{NE}}$.

Similarly as in Henry, Kitamura, and Salanié (2014)'s, notice that (15) can be written as:

$$p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}) = p\left(\mathbf{y}|\boldsymbol{\nu}^0\right) + \left[p\left(\mathbf{y}|\boldsymbol{\nu}^1\right) - p\left(\mathbf{y}|\boldsymbol{\nu}^0\right)\right]\Gamma(\mathbf{x}_{\mathrm{E}}). \tag{17}$$

By evaluating (17) at $\mathbf{x}_{\mathrm{E}}^0$ and $\mathbf{x}_{\mathrm{E}}^1$, it follows that:

$$p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right) = \left[p\left(\mathbf{y}|\boldsymbol{\nu}^1\right) - p\left(\mathbf{y}|\boldsymbol{\nu}^0\right)\right]\Upsilon \tag{18}$$

and

$$p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}) - p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right) = \left[p\left(\mathbf{y}|\boldsymbol{\nu}^1\right) - p\left(\mathbf{y}|\boldsymbol{\nu}^0\right)\right]\left[\Gamma(\mathbf{x}_{\mathrm{E}}) - \phi\right]. \tag{19}$$

By combining (18) and (19), one gets:

$$\Gamma(\mathbf{x}_{\mathrm{E}}) = \phi + \Upsilon\frac{p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}) - p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right)}{p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right)}. \tag{20}$$

By rearranging (17) evaluated at $\mathbf{x}_{\mathrm{E}}^0$ and using (18):

$$p\left(\mathbf{y}|\boldsymbol{\nu}^0\right) = p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right) - \frac{\phi}{\Upsilon}\left[p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right)\right]. \tag{21}$$

Finally, noting that $p(\mathbf{y}|\boldsymbol{\nu}^1) = p(\mathbf{y}|\boldsymbol{\nu}^0) + p(\mathbf{y}|\boldsymbol{\nu}^1) - p(\mathbf{y}|\boldsymbol{\nu}^0)$ and using (18) together with (21), it follows that:

$$p\left(\mathbf{y}|\boldsymbol{\nu}^1\right) = p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right) + \frac{1-\phi}{\Upsilon}\left[p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right)\right]. \tag{22}$$

The unobservable probabilities in (20), (21) and (22) are therefore functions of $\boldsymbol{\theta}$ and observable probabilities. Let's now turn to the conditions defining Θ_{I}. Consider the following functions:

$$L_0(\mathbf{y}) \equiv \frac{-p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}{p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}; \tag{23}$$

$$L_1(\mathbf{y}) \equiv \frac{1 - p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}{p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}; \tag{24}$$

$$Q(\mathbf{x}_{\mathrm{E}}) \equiv \frac{p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}) - p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}{p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}. \tag{25}$$

Moreover, let:

$$a_0 \equiv p\left(0,0|\mathbf{x}_{\mathrm{E}}^0\right) p\left(1,1|\mathbf{x}_{\mathrm{E}}^0\right) - p\left(1,0|\mathbf{x}_{\mathrm{E}}^0\right) p\left(0,1|\mathbf{x}_{\mathrm{E}}^0\right); \tag{26}$$

$$\begin{aligned} a_1 \equiv & p\left(0,0|\mathbf{x}_{\mathrm{E}}^0\right)\left[p\left(1,1|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(1,1|\mathbf{x}_{\mathrm{E}}^0\right)\right] + p\left(1,1|\mathbf{x}_{\mathrm{E}}^0\right)\left[p\left(0,0|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(0,0|\mathbf{x}_{\mathrm{E}}^0\right)\right] \\ & - p\left(1,0|\mathbf{x}_{\mathrm{E}}^0\right)\left[p\left(0,1|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(0,1|\mathbf{x}_{\mathrm{E}}^0\right)\right] - p\left(0,1|\mathbf{x}_{\mathrm{E}}^0\right)\left[p\left(1,0|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(1,0|\mathbf{x}_{\mathrm{E}}^0\right)\right]; \end{aligned} \tag{27}$$

$$\begin{aligned} a_2 \equiv & \left[p\left(0,0|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(0,0|\mathbf{x}_{\mathrm{E}}^0\right)\right]\left[p\left(1,1|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(1,1|\mathbf{x}_{\mathrm{E}}^0\right)\right] \\ & - \left[p\left(0,1|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(0,1|\mathbf{x}_{\mathrm{E}}^0\right)\right]\left[p\left(1,0|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(1,0|\mathbf{x}_{\mathrm{E}}^0\right)\right]. \end{aligned} \tag{28}$$

Without loss of generality, let $\Upsilon > 0$. Then, the identified set is defined in Proposition 1.

Proposition 1 (Identified set)**.** *Under Assumptions 1 to 4, $\boldsymbol{\theta} \in \Theta_I$ provided that:*

(i) $\max_{\mathbf{y}\in\mathcal{Y}^2} \{\min\{L_0(\mathbf{y}), L_1(\mathbf{y})\}\} \leq \frac{-\phi}{\Upsilon} < \min_{\mathbf{x}_E\in\mathcal{X}_E} \{Q(\mathbf{x}_E)\}$;

(ii) $\max_{\mathbf{x}_E\in\mathcal{X}_E} \{Q(\mathbf{x}_E)\} < \frac{1-\phi}{\Upsilon} \leq \min_{\mathbf{y}\in\mathcal{Y}^2} \{\max\{L_0(\mathbf{y}), L_1(\mathbf{y})\}\}$;

(iii) $a_2 \left[\frac{-\phi}{\Upsilon}\right]^2 - a_1 \frac{\phi}{\Upsilon} + a_0 = 0$;

(iv) $a_2 \left[\frac{1-\phi}{\Upsilon}\right]^2 + a_1 \frac{1-\phi}{\Upsilon} + a_0 = 0$.

Proof. See Appendix A.1. □

4.4 Properties of the identified set

I now turn to some interesting properties of Θ_{I} collected in two corollaries. Corollary 1 states that, in the simple $2 \times 2 \times 2$ case, if Θ_{I} is not empty, it is a singleton.

Corollary 1 (Singleton nonempty Θ_{I})**.** *If $\Theta_I \neq \emptyset$, then Θ_I is a singleton.*

Proof. See Appendix A.2. □

From Proposition 1 and Corollary 1, it is easy to represent the identified set graphically for the reparametrization $-\phi/\Upsilon$ and $(1-\phi)/\Upsilon$. In Figure 2, Θ_{I} is the singleton $(-\phi/\Upsilon)^{*}$, $((1-\phi)/\Upsilon)^{**}$. The shaded area corresponds to $\boldsymbol{\theta}$'s that satisfy conditions (i) and (ii) from Proposition 1. The couple $(-\phi/\Upsilon)^{*}$, $((1-\phi)/\Upsilon)^{**}$, if it exists, is the only one that satisfies conditions (iii) and (iv). As a result, $\Theta_{\mathrm{I}} \neq \emptyset$ if and only if this point falls in the shaded area, i.e.

$$\max_{\mathbf{y}\in\mathcal{Y}^2}\left\{\min\left\{L_0(\mathbf{y}), L_1(\mathbf{y})\right\}\right\} \leq \left(\frac{-\phi}{\Upsilon}\right)^{*} < \min_{\mathbf{x}_{\mathrm{E}}\in\mathcal{X}_{\mathrm{E}}}\left\{Q(\mathbf{x}_{\mathrm{E}})\right\}; \tag{29}$$

$$\max_{\mathbf{x}_{\mathrm{E}}\in\mathcal{X}_{\mathrm{E}}}\left\{Q(\mathbf{x}_{\mathrm{E}})\right\} < \left(\frac{1-\phi}{\Upsilon}\right)^{**} \leq \min_{\mathbf{y}\in\mathcal{Y}^2}\left\{\max\left\{L_0(\mathbf{y}), L_1(\mathbf{y})\right\}\right\}. \tag{30}$$

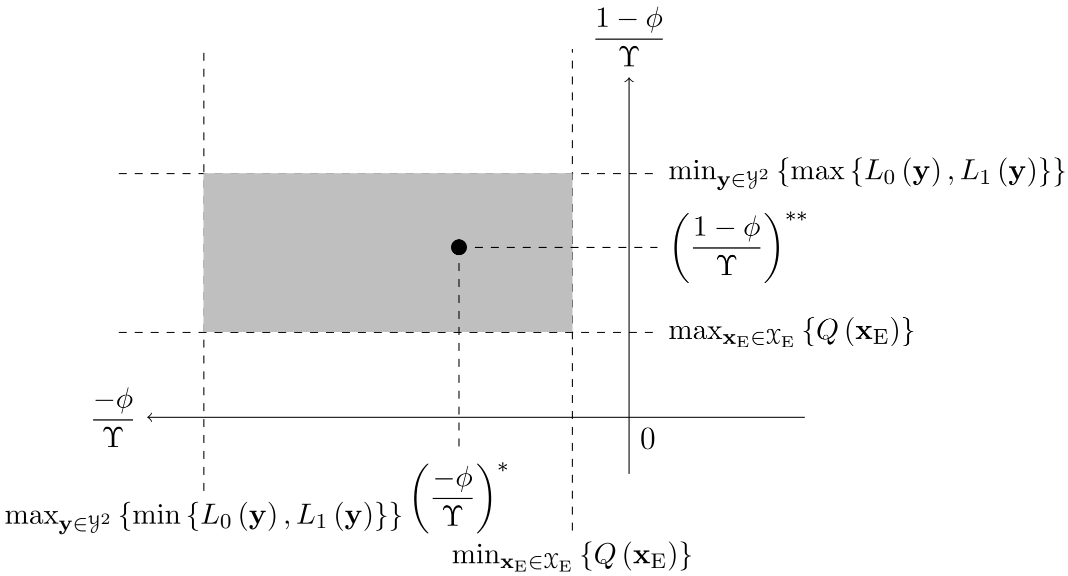

Figure 2: Graphical representation of the identified set for the $2 \times 2 \times 2$ case

A test based on the emptiness of Θ_{I} is sharp in the following sense. Given $\mathbf{x}_{\mathrm{NE}}$, if the true data generating process corresponds to a single equilibrium being realized at $\boldsymbol{\nu}^0$ and $\boldsymbol{\nu}^1$, then Θ_{I} is not empty. Conversely, if Θ_{I} is not empty, then the joint distribution of players' decisions conditional on $\mathbf{x}_{\mathrm{NE}}$ can be rationalized through a single equilibrium at $\boldsymbol{\nu}^0$ and $\boldsymbol{\nu}^1$. One drawback of considering testable implications is that the latter could hold even if the true data generating process does not correspond to a single equilibrium for $\boldsymbol{\nu}^0$ and/or $\boldsymbol{\nu}^1$. Nonetheless, a test based on $\Theta_{\mathrm{I}} \neq \emptyset$ is still informative since Θ_{I} being empty implies that the data cannot be rationalized by a single equilibrium and a finite mixture over common knowledge unobserved heterogeneity. In other words, $\Theta_{\mathrm{I}} \neq \emptyset$ is necessary, but not sufficient for equilibrium uniqueness. This limitation of the test can easily be understood from Figure 2: it is due to the fact that we do not observe the true $p(\mathbf{y}|\boldsymbol{\nu})$'s, but we know that they belong to a set. Some of the probabilities in this set may be rationalized through a single equilibrium even if the data is generated by multiple ones.

It is worth mentionning that the idea of checking testable implications or testing necessary conditions of a given assumption has also been suggested in other contexts. See for instance, Kitagawa (2015) for a test of instrument validity; Mourifié and Wan (2017) for local average treatment effect assumptions; Ghanem (2017) for identifying assumptions in nonseparable panel data models; Hsu, Liu, and Shi (2018) for generalized regression monotonicity; or Mourifié, Henry, and Méango (2018) for the Roy model. In many cases, the proposed test is implemented via a specification test, similarly as in the current setting.

Of course, given that we are representing the joint distribution of players' decisions as a double finite mixture, a natural question to ask is to which extent are we able to separate the two levels of mixing, i.e. the mixture over multiple equilibria and the mixture over common knowledge unobserved heterogeneity? Here, the identifying restrictions in Assumption 4 are key. In some sense, they allow one to identify $p(\mathbf{y}|\boldsymbol{\nu})$'s, regardless of the number of equilibria realized at each $\boldsymbol{\nu} \in \mathcal{V}$, using conditions (i) and (ii) in Proposition 1. Then, whether one can factor $p(\mathbf{y}|\boldsymbol{\nu})$ as the product of $p(y_1|\boldsymbol{\nu})$ and $p(y_2|\boldsymbol{\nu})$, i.e. conditions (iii) and (iv), determines whether one can rationalize the data through the single equilibrium assumption.

One important feature of Θ_{I}, as stated in Corollary 2 below, is that the emptiness of Θ_{I} does not depend on the weights associated with each realization of common knowledge unobserved heterogeneity. More precisely, consider two different data generating processes that are associated with the same $p(\mathbf{y}|\boldsymbol{\nu})$'s, but different $\Gamma(\mathbf{x}_{\mathrm{E}})$'s. Corollary 2 states that the identified sets corresponding to each data generating processes are either both empty or both nonempty.

Corollary 2 (Emptiness of Θ_{I} independent of $\Gamma(\mathbf{x}_{\mathrm{E}})$)**.** *Let $\bar{\boldsymbol{\theta}}^0 \neq \tilde{\boldsymbol{\theta}}^0$ correspond to two different data generating processes with choice probabilities such that $\bar{p}^0(\mathbf{y}|\boldsymbol{\nu}) = \tilde{p}^0(\mathbf{y}|\boldsymbol{\nu})$ $\forall \mathbf{y} \in \mathcal{Y}^2$ and $\forall \boldsymbol{\nu} \in \mathcal{V}$. Then, the corresponding identified sets are such that $\bar{\Theta}_I \neq \emptyset$ if and only if $\tilde{\Theta}_I \neq \emptyset$.*

Proof. See Appendix A.3. □

Corollary 2 highlights an important property of the identified set. For a given vector of equilibrium-specific choice probabilities, whether or not Θ_{I} is empty only depends on the equilibrium selection mechanism, not on the weights associated with each realization of the common knowledge unobserved heterogeneity. In that sense, a test based on $\Theta_{\mathrm{I}} \neq \emptyset$ effectively separates the two levels of mixing.

4.5 Identifying power of the empty identified set

As already mentioned above, a potential drawback of a test based on the testable implication $\Theta_{\mathrm{I}} \neq \emptyset$ is that it tests a necessary, but insufficient condition for equilibrium uniqueness. If there actually is a single equilibrium in the data generating process (and the joint distribution of players' decisions can be rationalized through a finite mixture over $\boldsymbol{\nu}^0$ and $\boldsymbol{\nu}^1$), then Θ_{I} is not empty. However, it could be the case that Θ_{I} is not empty even if the data has been generated by multiple equilibria. In that case, despite multiplicity of equilibria in the data generating process, the data could be rationalized by a single equilibrium. Therefore, for a test based on $\Theta_{\mathrm{I}} \neq \emptyset$ to be informative, it must be the case that Θ_{I} is empty at least for some data generating processes. There are two reasons why Θ_{I} may be empty:

1) *The data cannot be rationalized by a finite mixture over* $\boldsymbol{\nu} \in \mathcal{V}$. This would be the case if (i) and/or (ii) from Proposition 1 do not hold for any $\boldsymbol{\theta} \in \Theta$. More precisely, the data cannot be rationalized by such a finite mixture if either:

$$\min_{\mathbf{x}_{\mathrm{E}} \in \mathcal{X}_{\mathrm{E}}} \{Q(\mathbf{x}_{\mathrm{E}})\} \leq \max_{\mathbf{y} \in \mathcal{Y}^2} \{\min\{L_0(\mathbf{y}), L_1(\mathbf{y})\}\}; \tag{31}$$

$$\min_{\mathbf{y} \in \mathcal{Y}^2} \{\max\{L_0(\mathbf{y}), L_1(\mathbf{y})\}\} \leq \max_{\mathbf{x}_{\mathrm{E}} \in \mathcal{X}_{\mathrm{E}}} \{Q(\mathbf{x}_{\mathrm{E}})\}. \tag{32}$$

2) *The joint distribution of players' decisions conditional on* $\boldsymbol{\nu} \in \mathcal{V}$ *cannot be factored as the product of their marginal distributions.* In other words, either (iii) and/or (iv) from Proposition 1 do not hold. This could either be due to the quadratic equations not admitting a real solution or admitting solutions that fall outside of the bounds in (i) and (ii).

I now provide simulation evidence to show that Θ_{I} may indeed be empty in practice. I use the simple data generating process introduced in Example 1. Similarly as before, let $p_i(\boldsymbol{\nu})$ denote the probability that $y_i = 1$, given $\mathbf{x}_{\mathrm{NE}}$ and $\boldsymbol{\nu}$. Then, the BNE in pure strategies indexed by τ is such that:

$$\mathbf{p}^*(\boldsymbol{\nu}, \tau) \equiv \begin{bmatrix} p_1^*(\boldsymbol{\nu}, \tau) \\ p_2^*(\boldsymbol{\nu}, \tau) \end{bmatrix} = \begin{bmatrix} \Phi(x_{\mathrm{NE},1} + \nu_1 - 4p_2^*(\boldsymbol{\nu}, \tau)) \\ \Phi(x_{\mathrm{NE},2} + \nu_2 - 3p_1^*(\boldsymbol{\nu}, \tau)) \end{bmatrix}. \tag{33}$$

I consider two different data generating processes that take the same values of $\mathbf{x}_{\mathrm{NE}} \equiv [x_{\mathrm{NE},1}, x_{\mathrm{NE},2}]' = [1,1]'$, but vary according to the values of $\boldsymbol{\nu}^0$ and $\boldsymbol{\nu}^1$. In the first one, let $\boldsymbol{\nu}^0 = [1, 0.5]'$ and $\boldsymbol{\nu}^1 = [1.25, 1]'$. This case admits three solutions for each $\boldsymbol{\nu}$, with two of these three equilibria being stable. Even though stability is not required for the identification result to hold, I focus on stable equilibria since they have a more natural economic interpretation. These equilibria are such that $\mathbf{p}^*(\boldsymbol{\nu}^0, \tau = 1) = [0.0499, 0.9116]'$, $\mathbf{p}^*(\boldsymbol{\nu}^0, \tau = 2) = [0.9501, 0.0884]'$ for $\boldsymbol{\nu}^0$; $\mathbf{p}^*(\boldsymbol{\nu}^1, \tau = 1) = [0.0527, 0.9672]'$, $\mathbf{p}^*(\boldsymbol{\nu}^1, \tau = 2) = [0.9003, 0.2417]'$ for $\boldsymbol{\nu}^1$.

Figure 3 summarizes whether $\Theta_{\mathrm{I}} \neq \emptyset$ for $\lambda^*(\tau|\boldsymbol{\nu}) \in \{0, 0.05, \ldots, 0.95, 1\}$ with the vertical axis corresponding to $\lambda^*(\tau = 1|\boldsymbol{\nu}^0)$ and the horizontal axis corresponding to $\lambda^*(\tau = 1|\boldsymbol{\nu}^1)$. While Θ_{I} is empty in the white area, it is nonempty in the black ones. For the couples $(\lambda^*(\tau = 1|\boldsymbol{\nu}^0), \lambda^*(\tau = 1|\boldsymbol{\nu}^1)) \in \{(0,0), (0,1), (1,0), (1,1)\}$, i.e. the four "corners", we have $\Theta_{\mathrm{I}} \neq \emptyset$. This observation is as expected: these four equilibrium selection mechanisms are the only ones which satisfy the single equilibrium in the data assumption. The fact that other selection mechanisms are associated with a non-empty identified set confirms that the data may sometimes be rationalized by a single equilibrium, even if it is generated by multiple equilibria. If the whole area was black, a test based on $\Theta_{\mathrm{I}} \neq \emptyset$ would not be informative. One could then always rationalize these mixtures over two equilibria by a single equilibrium. The presence of the white area implies that for some equilibrium selection mechanisms in the data generating process, the resulting joint distributions of players' decisions simply cannot be rationalized by a single equilibrium.

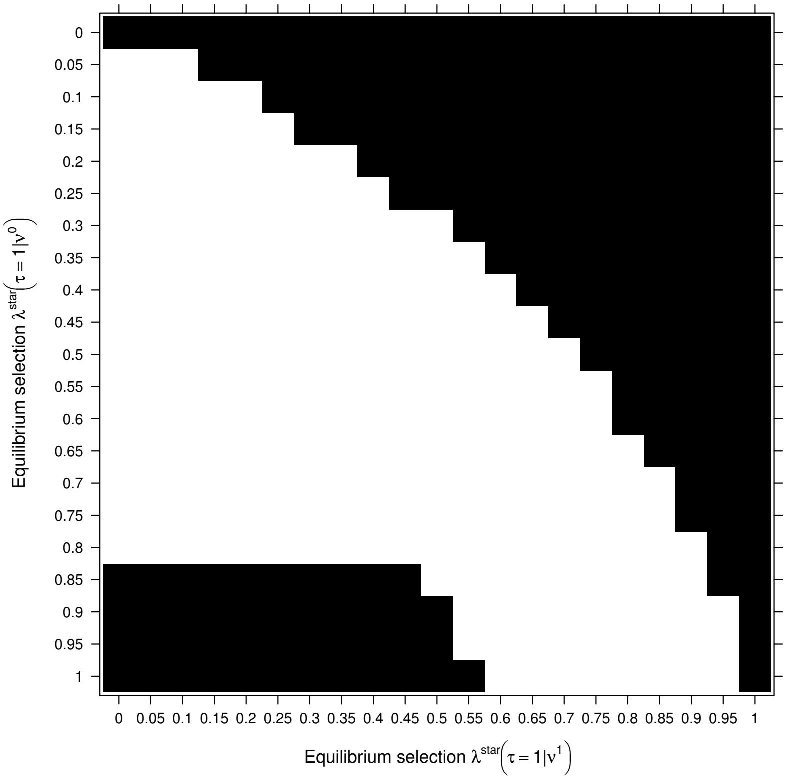

Figure 3: Emptiness of $\Theta_{\rm I}$ – Case 1

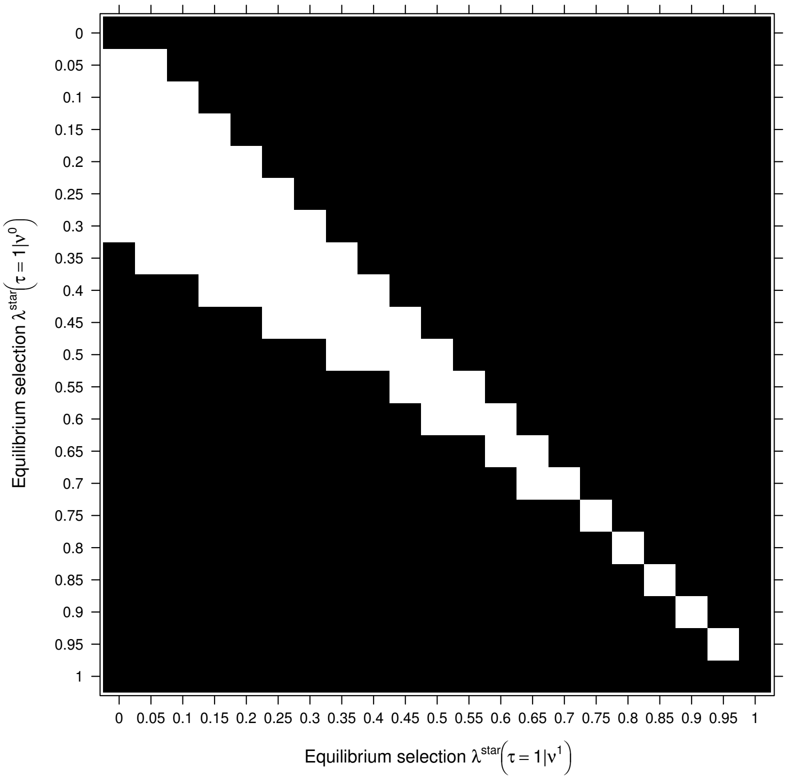

Figure 4: Emptiness of $\Theta_{\rm I}$ – Case 2

The second data generating process is such that $\boldsymbol{\nu}^0 = [1.30, 1]'$ and $\boldsymbol{\nu}^1 = [1.25, 1]'$. In this case, the equilibria kept are $\mathbf{p}^* (\boldsymbol{\nu}^0, \tau = 1) = [0.0590, 0.9659]'$, $\mathbf{p}^* (\boldsymbol{\nu}^0, \tau = 2) = [0.9201, 0.2235]'$ for $\boldsymbol{\nu}^0$; $\mathbf{p}^* (\boldsymbol{\nu}^1, \tau = 1) = [0.0527, 0.9672]'$, $\mathbf{p}^* (\boldsymbol{\nu}^1, \tau = 2) = [0.9003, 0.2417]'$ for $\boldsymbol{\nu}^1$. Figure 4 suggests that Θ_{I} is nonempty for most equilibrium selection mechanisms. Notice that an important difference between this data generating process and the previous one is that equilibria now barely vary across $\boldsymbol{\nu}$'s. Nonetheless, the testable implications are still informative.

5 Statistical test

5.1 A two-stage minimum distance approach

The identification result derived above suggests that Θ_{I} being nonempty is an informative necessary condition of the single equilibrium assumption in the presence of discrete common knowledge unobserved heterogeneity. More formally, this condition can be tested as:

$$\mathrm{H}_0 : \Theta_{\mathrm{I}} \neq \emptyset; \quad \mathrm{H}_1 : \Theta_{\mathrm{I}} = \emptyset. \tag{34}$$

In other words, the proposed test of the single equilibrium in the data assumption boils down to a specification test in partial identification.

There are three important features of the problem at hand that should be taken into account when thinking about the inference method to be used to test this null hypothesis. First, the conditions defining Θ_{I} are functions of some estimable probabilities, which are expectations of discrete variables. However, since these expectations enter the conditions nonlinearly, the usual moment inequalities framework does not readily apply. Second, these probabilities are unknown to the researcher and must therefore be consistently estimated prior to constructing the identified set. Third, some of the conditions defining Θ_{I} are equality constraints.

The two-stage minimum distance estimator proposed by Shi and Shum (2015) allows us to take advantage of these three features. Their approach can be interpreted as an extension of classical minimum distance estimation procedures to include inequality constraints and partial identification. In order to apply their method, one must be able to write the conditions defining Θ_{I} as:

1. Some equality constraints that depend on first-stage probabilities, the parameters of interest and potentially some nuisance parameters.

2. Some inequality constraints that depend on the parameters of interest and potentially some nuisance parameters, but not the first-stage probabilities.

Let $\mathbf{p}$ denote the vector of joint probabilities of players' decisions $\forall \mathbf{y} \in \mathcal{Y}^2$ and $\forall \mathbf{x}_{\mathrm{E}} \in \mathcal{X}_{\mathrm{E}}$. Let $\hat{\mathbf{p}}$ be the first-stage estimate of the population probabilities $\mathbf{p}^0$ such that:

$$M^{1/2} \left(\hat{\mathbf{p}} - \mathbf{p}^0 \right) \xrightarrow{d} N \left(\mathbf{0}, \boldsymbol{\Sigma} \right) \tag{35}$$

where $\mathbf{\Sigma}$ is a block-diagonal matrix with $\mathbf{x}_{\mathrm{E}}$-specific blocks $\left(M/M\left(\mathbf{x}_{\mathrm{E}}\right)\right)\mathbf{\Sigma}\left(\mathbf{x}_{\mathrm{E}}\right)$ where $M\left(\mathbf{x}_{\mathrm{E}}\right)$ is the number of markets with realization $\mathbf{x}_{\mathrm{E}}$. Moreover, let $\boldsymbol{\eta} \in \mathcal{H}$ be a vector of nuisance parameters to be defined below. Essentially, these nuisance parameters are used to transform inequalities that depend on $\mathbf{p}$ into equalities.

There are two types of equality conditions in this context. First, let the vector $\mathbf{g}^{\mathrm{e}}\left(\mathbf{p},\boldsymbol{\theta}\right)$ be some equality conditions such that $\mathbf{g}^{\mathrm{e}}\left(\mathbf{p},\boldsymbol{\theta}\right) = \mathbf{0}$. These are conditions defining the identified set that do not need the introduction of nuisance parameters to be written as equalities, which is the reason why they do not depend on $\boldsymbol{\eta}$. Second, let $\mathbf{g}^{\mathrm{ie},1}\left(\mathbf{p},\boldsymbol{\theta}\right) - \boldsymbol{\eta}$ be a vector of equality conditions such that $\mathbf{g}^{\mathrm{ie},1}\left(\mathbf{p},\boldsymbol{\theta}\right) - \boldsymbol{\eta} = \mathbf{0}$. These conditions correspond to inequalities that have been transformed into equalities by introducing nuisance parameters. Finally, one must introduce the remaining inequality conditions. Let $\mathbf{g}^{\mathrm{ie},2}\left(\boldsymbol{\theta},\boldsymbol{\eta}\right)$, which do not depend on $\mathbf{p}$, be such that $\mathbf{g}^{\mathrm{ie},2}\left(\boldsymbol{\theta},\boldsymbol{\eta}\right) \geq \mathbf{0}$. Typically, one must rewrite all equalities such that the matrix of their joint variance (defined below) is full rank. The details for the $2 \times 2 \times 2$ case are provided in Appendix B.

At a given $\boldsymbol{\theta}$ the full vector of equality conditions evaluated at the first stage estimates $\hat{\mathbf{p}}$, i.e. $\mathbf{g}\left(\hat{\mathbf{p}},\boldsymbol{\theta},\boldsymbol{\eta}\right) \equiv \left[\mathbf{g}^{\mathrm{e}}\left(\hat{\mathbf{p}},\boldsymbol{\theta}\right)', \mathbf{g}^{\mathrm{ie},1}\left(\hat{\mathbf{p}},\boldsymbol{\theta}\right)' - \boldsymbol{\eta}'\right]'$, inherits the asymptotic normality of $\hat{\mathbf{p}}$:

$$M^{1/2}\left(\mathbf{g}\left(\hat{\mathbf{p}},\boldsymbol{\theta},\boldsymbol{\eta}\right) - \mathbf{g}\left(\mathbf{p}^{0},\boldsymbol{\theta},\boldsymbol{\eta}\right)\right) \xrightarrow{d} N\left(\mathbf{0},\mathbf{W}\left(\boldsymbol{\theta}\right)\right) \tag{36}$$

where $\mathbf{W}\left(\boldsymbol{\theta}\right) \equiv \left[\mathbf{G}\left(\mathbf{p}^{0},\boldsymbol{\theta}\right)\mathbf{\Sigma}\mathbf{G}\left(\mathbf{p}^{0},\boldsymbol{\theta}\right)'\right]$ and $\mathbf{G}\left(\mathbf{p}^{0},\boldsymbol{\theta}\right)$ is the Jacobian of $\mathbf{g}\left(\mathbf{p},\boldsymbol{\theta},\boldsymbol{\eta}\right)$ with respect to $\mathbf{p}$, evaluated at $\mathbf{p}^{0}$.

Consider the following profiled criterion function:

$$Q\left(\hat{\mathbf{p}},\boldsymbol{\theta}\right) = \min_{\boldsymbol{\eta}\in\mathcal{H}(\boldsymbol{\theta})} \begin{bmatrix} \mathbf{g}^{\mathrm{e}}\left(\hat{\mathbf{p}},\boldsymbol{\theta}\right) \\ \mathbf{g}^{\mathrm{ie},1}\left(\hat{\mathbf{p}},\boldsymbol{\theta}\right) - \boldsymbol{\eta} \end{bmatrix}' \hat{\mathbf{W}}\left(\boldsymbol{\theta}\right)^{-1} \begin{bmatrix} \mathbf{g}^{\mathrm{e}}\left(\hat{\mathbf{p}},\boldsymbol{\theta}\right) \\ \mathbf{g}^{\mathrm{ie},1}\left(\hat{\mathbf{p}},\boldsymbol{\theta}\right) - \boldsymbol{\eta} \end{bmatrix} \tag{37}$$

where $\mathcal{H}\left(\boldsymbol{\theta}\right) \equiv \left\{\boldsymbol{\eta} \in \mathcal{H} : \mathbf{g}^{\mathrm{ie},2}\left(\boldsymbol{\theta},\boldsymbol{\eta}\right) \geq \mathbf{0}\right\}$ and $\hat{\mathbf{W}}\left(\boldsymbol{\theta}\right)$ is a consistent estimate of $\mathbf{W}\left(\boldsymbol{\theta}\right)$ obtained by replacing $\hat{\mathbf{p}}$ with $\mathbf{p}^{0}$. Given the quadratic form of this criterion function, one can define the identified set as:

$$\Theta_{\mathrm{I}} = \left\{\boldsymbol{\theta} \in \Theta : Q\left(\mathbf{p}^{0},\boldsymbol{\theta}\right) = 0\right\}. \tag{38}$$

Testing the emptiness of Θ_{I} can be done by checking whether the confidence set, i.e. the collection of $\boldsymbol{\theta}$'s that belong to Θ_{I} with confidence level at least $1-\alpha$, is empty.[10] Let the confidence set be defined as:

$$\mathrm{CS}_{M}\left(\alpha\right) = \left\{\boldsymbol{\theta} \in \Theta : MQ\left(\hat{\mathbf{p}},\boldsymbol{\theta}\right) \leq c_{M}\left(\boldsymbol{\theta}, 1-\alpha\right)\right\} \tag{39}$$

where $c_{M}\left(\boldsymbol{\theta}, 1-\alpha\right)$ is the $1-\alpha$-th quantile of $MQ\left(\hat{\mathbf{p}},\boldsymbol{\theta}\right)$'s distribution. One can take advantage of the asymptotic normality of $\mathbf{g}\left(\hat{\mathbf{p}},\boldsymbol{\theta},\boldsymbol{\eta}\right)$ when computing $c_{M}\left(\boldsymbol{\theta}, 1-\alpha\right)$. Shi and Shum (2015)

[10] Bugni, Canay, and Shi (2015) refer to such a specification test as the by-product test. They propose alternative approaches that typically have relatively better power. However, their tests are designed for identified sets defined according to moment inequalities and may not directly apply to the current setting.

propose to use the corresponding quantile of the variable $J_M(\boldsymbol{\theta})$ defined as:

$$J_M(\boldsymbol{\theta}) = \min_{\mathbf{h}\in\mathcal{H}(\boldsymbol{\theta})-\hat{\boldsymbol{\eta}}(\boldsymbol{\theta})} \begin{bmatrix} \mathbf{z}_M^{\mathrm{e}} \\ \mathbf{z}_M^{\mathrm{ie},1} - \kappa_M^{-1} M^{1/2}\mathbf{h} \end{bmatrix}' \hat{\mathbf{W}}(\boldsymbol{\theta})^{-1} \begin{bmatrix} \mathbf{z}_M^{\mathrm{e}} \\ \mathbf{z}_M^{\mathrm{ie},1} - \kappa_M^{-1} M^{1/2}\mathbf{h} \end{bmatrix} \tag{40}$$

where $\left[\mathbf{z}_M^{\mathrm{e}\prime}, \mathbf{z}_M^{\mathrm{ie},1\prime}\right]' \sim N\left(\mathbf{0}, \hat{\mathbf{W}}(\boldsymbol{\theta})\right)$; $\{\kappa_M\}$ is a sequence of tuning parameters that diverges to infinity; and

$$\hat{\boldsymbol{\eta}}(\boldsymbol{\theta}) = \min_{\boldsymbol{\eta}\in\mathcal{H}(\boldsymbol{\theta})} \left[\mathbf{g}^{\mathrm{ie},1}(\hat{\mathbf{p}}, \boldsymbol{\theta}) - \boldsymbol{\eta}\right]' \left[\mathbf{g}^{\mathrm{ie},1}(\hat{\mathbf{p}}, \boldsymbol{\theta}) - \boldsymbol{\eta}\right]. \tag{41}$$

Consider the following non-randomized decision rule:

$$\xi_M = \begin{cases} 1, & \text{if } \mathrm{CS}_M(\alpha) = \emptyset \\ 0, & \text{otherwise.} \end{cases} \tag{42}$$

Let $\mathrm{E}_{\boldsymbol{\theta}}[\cdot]$ denote the expectation under the data generating process corresponding to $\boldsymbol{\theta}$. In particular, for $\boldsymbol{\theta} \in \Theta_{\mathrm{I}}$, $\mathrm{E}_{\boldsymbol{\theta}}[\xi_M]$ is the probability of rejecting $\mathrm{H}_0 : \Theta_{\mathrm{I}} \neq \emptyset$ when it's true. Proposition 2 states that the statistical test of this null hypothesis based on the decision rule ξ_M is asymptotically level α.

Proposition 2 (Asymptotic level of the test)**.** *Suppose that: (i) $\mathbf{G}(\mathbf{p}, \boldsymbol{\theta})$ is well defined and continuous in $\boldsymbol{\theta}$ and $\mathbf{p}$; (ii) $\hat{\mathbf{W}}(\boldsymbol{\theta}) \xrightarrow{p} \mathbf{W}(\boldsymbol{\theta})$; (iii) $\mathbf{W}(\boldsymbol{\theta})$ is invertible $\forall \boldsymbol{\theta} \in \Theta$; (iv) Θ is compact; (v) $\mathbf{g}(\mathbf{p}, \boldsymbol{\theta}, \boldsymbol{\eta})$ is continuous in $\boldsymbol{\theta}$ for all $\mathbf{p}$; (vi) $\mathbf{g}^{ie,2}(\boldsymbol{\theta}, \boldsymbol{\eta})$ is continuous in $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$; and (vii) $\mathcal{H}(\boldsymbol{\theta})$ is convex $\forall \boldsymbol{\theta} \in \Theta_I$. Then:*

$$\limsup_{M\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_I} \mathrm{E}_{\boldsymbol{\theta}}[\xi_M] \leq \alpha.$$

Proof. The result follows from Shi and Shum (2015, Theorem 4.1, p. 503). Details are provided in Appendix A.4. □

5.2 Monte-Carlo simulations

I now provide simulation evidence in order to investigate the statistical size and power of the test based on ξ_M. More precisely, I investigate the probability of rejecting the null hypothesis $\Theta_{\mathrm{I}} \neq \emptyset$ for equilibrium selection mechanisms that correspond to a single equilibrium and to multiple equilibria.

Once again, I consider the same data generating process as in the simple game of market entry introduced in Example 1. Let the values of $\mathbf{x}_{\mathrm{NE}}$, $\boldsymbol{\nu}^0$ and $\boldsymbol{\nu}^1$ be as in the first data generating process used when investigating the identifying power of the testable restriction in Section 4.5 and illustrated in Figure 3. I consider six different realizations of $\mathbf{x}_{\mathrm{E}}$. For each of these realizations, the weights associated with $\boldsymbol{\nu}^1$ are listed in Table 1. Since $\Gamma(\mathbf{x}_{\mathrm{E}}^0) = 0.25$ and $\Gamma(\mathbf{x}_{\mathrm{E}}^1) = 0.8$, it follows that $\phi^0 = 0.25$ and $\Upsilon^0 = 0.55$ in the data generating process.

Different data generating processes are created by varying equilibrium selection mechanisms for $\boldsymbol{\nu}^0$ and $\boldsymbol{\nu}^1$. I consider six different cases labelled A to F. Cases A to D (which are represented

Table 1: Mixture weights for unobserved heterogeneity

	$\mathbf{x}_{\mathrm{E}}^0$	$\mathbf{x}_{\mathrm{E}}^1$	$\mathbf{x}_{\mathrm{E}}^2$	$\mathbf{x}_{\mathrm{E}}^3$	$\mathbf{x}_{\mathrm{E}}^4$	$\mathbf{x}_{\mathrm{E}}^5$
$\mathbf{\Gamma}(\mathbf{x}_E)$	0.25	0.8	0.9	0.15	0.5	0.7

Note: $\mathbf{\Gamma}(\mathbf{x}_E)$ is the weight associated with $\boldsymbol{\nu}^1$, conditional on $\mathbf{x}_E$.

by the four corners in Figure 3) satisfy the single equilibrium in the data assumption and, therefore, the corresponding true identified sets Θ_{I} are nonempty. Cases E and F mix two equilibria for both $\boldsymbol{\nu}^0$ and $\boldsymbol{\nu}^1$. Both cases therefore violate the single equilibrium in the data assumption. However, the equilibrium selection mechanisms are such that Θ_{I} is empty for case E, but nonempty for case F. In other words, while the data are generated from multiple equilibria in both cases, the data generating process in case F can still be rationalized by a single equilibrium. As already mentioned above, this feature of the proposed test is a consequence of testing necessary conditions that are not sufficient for equilibrium uniqueness.

For a given Monte Carlo sample, the joint conditional choice probabilities in $\mathbf{p}^0$ are estimated using a simple frequency count estimator. In order to compute ξ_M, one must span the parameter space Θ, at least until $\mathrm{CS}_M(\alpha)$ is found to be nonempty. I consider the grid $\{0.025, 0.05, \ldots, 0.95, 0.975\}$ for both ϕ and Υ. For each possible value of $\boldsymbol{\theta}$ in this discretized parameter space (subject to $\phi + \Upsilon < 1$), I compute $MQ(\hat{\mathbf{p}}, \boldsymbol{\theta})$ and the quantile $c_M(\boldsymbol{\theta}, 1-\alpha)$. When computing this quantile, I consider 100 draws of the random variable $J_M(\boldsymbol{\theta})$ defined in (40) and, following Shi and Shum (2015)'s advice, I set $\kappa_M = \sqrt{\ln[M]}$. For each data generating process, I consider two different sample sizes, $M \in \{600, 1200\}$, which are equally split among realized $\mathbf{x}_E$'s. The rejection probabilities of the test, for $\alpha = 0.1$ and $\alpha = 0.05$ are computed using 100 Monte Carlo samples.[11]

Table 2: Rejection probabilities

Cases	$\lambda^*(\tau = 1 \mid \boldsymbol{\nu}^0)$	$\lambda^*(\tau = 1 \mid \boldsymbol{\nu}^1)$	Empty Θ_{I}	$M = 600$		$M = 1200$	
				$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$
A	0	0	No	0.07	0.00	0.04	0.00
B	0	1	No	0.01	0.01	0.00	0.00
C	1	0	No	0.01	0.00	0.00	0.00
D	1	1	No	0.06	0.00	0.11	0.00
E	0.45	0.30	Yes	0.63	0.62	0.71	0.69
F	0.2	0.85	No	0.01	0.01	0.03	0.03

Notes: Rejection probabilities computed over 100 Monte Carlo samples. Critical values used to construct the confidence set correspond to the $1-\alpha$-th quantile obtained from 100 draws of $J_M(\boldsymbol{\theta})$. The column "Empty Θ_{I}" indicates if the true identified set in the data generating process is empty.

The probabilities of rejecting the null hypothesis that $\Theta_{\mathrm{I}} \neq \emptyset$ are reported in Table 2. A

[11] I dropped samples leading to some estimated joint choice probabilities equal to zero or such that $\hat{\mathbf{p}}(0, 0 \mid \mathbf{x}_E^1) - \hat{\mathbf{p}}(0, 0 \mid \mathbf{x}_E^0) = 0$ since these samples led to ill-defined test statistics.

few comments are worth making. First, cases A to D confirm that the test defined by ξ_M in (42) is level α even in relatively small samples. However, the test is somewhat conservative. This observation is in fact a common feature of specification tests in partial identification based on checking whether the confidence set is empty. In the context of identified sets defined according to moment inequalities, Andrews and Guggenberger (2009) and Andrews and Soares (2010) have already pointed out that such tests may have asymptotic size strictly smaller than α.

The rejection probabilities reported in Table 2 are also useful to assess the power of the test. Cases E and F both correspond to samples generated from multiple equilibria. However, only case E is associated with an empty identified set in the data generating process. This is the reason why the proposed test leads to relatively high rejection probabilities which are growing with sample size in case E, but low rejection rates in case F. To the extent that the test proposed here is a test of the emptiness of Θ_{I}, i.e. a sharp test of necessary conditions for equilibrium uniqueness, the rejection probabilities reported in Table 2 are as expected. The fact that the power of the test is large for some alternatives implies that this test is still statistically informative.

6 Concluding remarks

To sum up, the test for equilibrium uniqueness that I present addresses two important issues associated with the procedures previously proposed in the literature. First, I allow for common knowledge payoff-relevant unobservables, therefore making the information structure of the model relatively more flexible. Second, the test that I propose does not require the estimation of payoff functions to separate the problems of multiple equilibria and common knowledge unobserved heterogeneity. The latter feature of the test is of interest for empirical researchers interested in testing equilibrium uniqueness in hope of leveraging multiple equilibria as a source of variation when commonly-used exclusion restrictions are not available. Moreover, no parametric assumption is needed for the payoff functions nor the distributions of the unobservables, besides the finite mixture representation. The main identifying assumption is the existence of an observable variable which can be interpreted as a proxy variable for the common knowledge unobserved heterogeneity. The test boils down to a specification test and it can be implemented through a two-stage minimum distance procedure that has nice properties verified through simulations.

There are at least two natural extensions of the current test that I leave for future research. First, it may be interesting to allow for continuous common knowledge unobservables. In this vein, it may be possible to use a deconvolution argument along the lines of Khan and Nekipelov (2018). However, the results available typically impose more structure on the payoffs, the unobservables and the equilibrium selection mechanism compared to the test proposed above.

Second, the simulation results presented in Sections 4.5 and 5.2 clearly suggest that the test has higher power under some alternatives than others. For now, I do not have a clear characterization of the directions for which the test has more power. This is a feature of the test that is worth investigating.

Appendix A Proofs

A.1 Proof of Proposition 1

Conditions (i) and (ii) are an application of Henry, Kitamura, and Salanié (2014, Section 2)'s results to the current setting. They come from $p(\mathbf{y}|\boldsymbol{\nu})$'s and $\Gamma(\mathbf{x}_{\mathrm{E}})$'s being in the unit interval. First, consider the case where $p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0) > 0$. From equations (20), (21) and (22):

$$0 \leq p\left(\mathbf{y}|\boldsymbol{\nu}^0\right) \leq 1 \Rightarrow L_0(\mathbf{y}) \leq \frac{-\phi}{\Upsilon} \leq L_1(\mathbf{y}); \tag{43}$$

$$0 \leq p\left(\mathbf{y}|\boldsymbol{\nu}^1\right) \leq 1 \Rightarrow L_0(\mathbf{y}) \leq \frac{1-\phi}{\Upsilon} \leq L_1(\mathbf{y}); \tag{44}$$

$$\Gamma(\mathbf{x}_{\mathrm{E}}) > 0 \Rightarrow \frac{-\phi}{\Upsilon} < Q(\mathbf{x}_{\mathrm{E}}); \tag{45}$$

$$\Gamma(\mathbf{x}_{\mathrm{E}}) < 1 \Rightarrow \frac{1-\phi}{\Upsilon} > Q(\mathbf{x}_{\mathrm{E}}). \tag{46}$$

Notice that $p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0) > 0$ implies $L_0(\mathbf{y}) \leq Q(\mathbf{x}_{\mathrm{E}}) \leq L_1(\mathbf{y})$. Therefore the upper bound on $-\phi/\Upsilon$ in (43) is satisfied whenever (45) holds. Similarly, the lower bound on $(1-\phi)/\Upsilon$ in (44) is satisfied whenever (46) holds. Inequalities (43) to (46) can therefore be summarized as:

$$L_0(\mathbf{y}) \leq \frac{-\phi}{\Upsilon} < Q(\mathbf{x}_{\mathrm{E}}); \tag{47}$$

$$Q(\mathbf{x}_{\mathrm{E}}) < \frac{1-\phi}{\Upsilon} \leq L_1(\mathbf{y}). \tag{48}$$

A similar argument for the case where $p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - p(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0) < 0$, which in turn implies that $L_1(\mathbf{y}) \leq Q(\mathbf{x}_{\mathrm{E}}) \leq L_0(\mathbf{y})$, gives:

$$L_1(\mathbf{y}) \leq \frac{-\phi}{\Upsilon} < Q(\mathbf{x}_{\mathrm{E}}); \tag{49}$$

$$Q(\mathbf{x}_{\mathrm{E}}) < \frac{1-\phi}{\Upsilon} \leq L_0(\mathbf{y}). \tag{50}$$

Since inequalities (47) to (50) must hold for all $\mathbf{y} \in \mathcal{Y}^2$ and all $\mathbf{x}_{\mathrm{E}} \in \mathcal{X}_{\mathrm{E}}$, the restrictions on $\boldsymbol{\theta}$ coming from $p(\mathbf{y}|\boldsymbol{\nu})$'s and $\Gamma(\mathbf{x}_{\mathrm{E}})$'s being in the unit interval boil down to:

$$\max_{\mathbf{y}\in\mathcal{Y}^2}\{\min\{L_0(\mathbf{y}), L_1(\mathbf{y})\}\} \leq \frac{-\phi}{\Upsilon} < \min_{\mathbf{x}_{\mathrm{E}}\in\mathcal{X}_{\mathrm{E}}}\{Q(\mathbf{x}_{\mathrm{E}})\}; \tag{51}$$

$$\max_{\mathbf{x}_{\mathrm{E}}\in\mathcal{X}_{\mathrm{E}}}\{Q(\mathbf{x}_{\mathrm{E}})\} < \frac{1-\phi}{\Upsilon} \leq \min_{\mathbf{y}\in\mathcal{Y}^2}\{\max\{L_0(\mathbf{y}), L_1(\mathbf{y})\}\} \tag{52}$$

which correspond to conditions (i) and (ii). Conditions (iii) and (iv) hold if the single equilibrium in the data assumption holds, i.e. $|\mathcal{T}^*(\boldsymbol{\nu}^0)| = |\mathcal{T}^*(\boldsymbol{\nu}^1)| = 1$. More precisely, these are restrictions on $\boldsymbol{\theta}$ that hold if players' decisions are independent conditional on $\mathbf{x}_{\text{NE}}$ and $\boldsymbol{\nu}$. This independence is a consequence of Assumption 1(i), i.e. the independence of private information shocks across players. Following Kasahara and Shimotsu (2014), one way to characterize this independence is by evaluating the rank of the following matrix $\forall \boldsymbol{\nu} \in \mathcal{V}$:

$$\mathbf{P}(\boldsymbol{\nu}) \equiv \begin{bmatrix} p(0,0|\boldsymbol{\nu}) & p(0,1|\boldsymbol{\nu}) \\ p(1,0|\boldsymbol{\nu}) & p(1,1|\boldsymbol{\nu}) \end{bmatrix}. \tag{53}$$

Then, $p(\mathbf{y}|\boldsymbol{\nu}) = p(y_1|\boldsymbol{\nu})\, p(y_2|\boldsymbol{\nu})$ and $\text{rank}\{\mathbf{P}(\boldsymbol{\nu})\} = 1$, which can be written as:

$$p(0,0|\boldsymbol{\nu})\, p(1,1|\boldsymbol{\nu}) - p(1,0|\boldsymbol{\nu})\, p(0,1|\boldsymbol{\nu}) = 0. \tag{54}$$

Evaluating (54) at each $\boldsymbol{\nu} \in \mathcal{V}$ using (21) and (22) generates quadratic equations in $-\phi/\Upsilon$ and $(1-\phi)/\Upsilon$ that correspond to conditions (iii) and (iv) defining Θ_{I}.

A.2 Proof of Corollary 1

This statement follows from the quadratic equations associated with conditions (iii) and (iv) in Proposition 1 having at most one solution in $-\phi/\Upsilon$ and $(1-\phi)/\Upsilon$ after having fixed the sign of Υ. To see this, notice that these quadratic equations each have at most two solutions. Denote these solutions $(-\phi/\Upsilon)^*$ and $(-\phi/\Upsilon)^{**}$ for condition (iii); $((1-\phi)/\Upsilon)^*$ and $((1-\phi)/\Upsilon)^{**}$ for condition (iv). There are therefore four possible combinations of solutions. Notice that, since the quadratic equations in (iii) and (iv) have the same coefficients a_0, a_1 and a_2, the solutions are the same for both equations, i.e. $(-\phi/\Upsilon)^* = ((1-\phi)/\Upsilon)^*$ and $(-\phi/\Upsilon)^{**} = ((1-\phi)/\Upsilon)^{**}$. However, since $-\phi/\Upsilon \neq (1-\phi)/\Upsilon$, there are two potentially valid solution couples: $(-\phi/\Upsilon)^*$, $((1-\phi)/\Upsilon)^{**}$ and $(-\phi/\Upsilon)^{**}$, $((1-\phi)/\Upsilon)^*$. Finally, since by definition $\phi \in (0,1)$, $-\phi/\Upsilon$ and $(1-\phi)/\Upsilon$ must be of opposite signs. Once one fixes $\Upsilon > 0$, which is without loss of generality, there is only one solution couple such that $-\phi/\Upsilon < 0$ and $(1-\phi)/\Upsilon > 0$. If it exists, this solution couple defines a system of two equations in two unknowns which corresponds to at most one $\boldsymbol{\theta} \in \Theta_{\text{I}}$.

A.3 Proof of Corollary 2

Let $\bar{p}(\mathbf{y}|\boldsymbol{\nu})$ be the choice probability constructed from (21) and (22) using an arbitrary $\bar{\boldsymbol{\theta}} \in \bar{\Theta}_{\text{I}}$. Similarly for $\tilde{p}(\mathbf{y}|\boldsymbol{\nu})$. In order to show the result stated in the Corollary 2, it is useful to show an intermediary one. I first show that, if $\bar{\Theta}_{\text{I}} \neq \emptyset$ and $\tilde{\Theta}_{\text{I}} \neq \emptyset$, then $\bar{p}(\mathbf{y}|\boldsymbol{\nu}) = \tilde{p}(\mathbf{y}|\boldsymbol{\nu})$ $\forall \bar{\boldsymbol{\theta}} \in \bar{\Theta}_{\text{I}}$, $\forall \tilde{\boldsymbol{\theta}} \in \tilde{\Theta}_{\text{I}}$, $\forall \mathbf{y} \in \mathcal{Y}^2$ and $\forall \boldsymbol{\nu} \in \mathcal{V}$.

This intermediary result is straightforward to show if $\bar{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}}^0 \in \bar{\Theta}_{\text{I}}$ and $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}^0 \in \tilde{\Theta}_{\text{I}}$, i.e. when the true data generating processes are included in the identified sets. The result then holds by construction since $\bar{p}^0(\mathbf{y}|\boldsymbol{\nu}) = \tilde{p}^0(\mathbf{y}|\boldsymbol{\nu})$ and it holds for all elements of the identified sets since $\bar{\Theta}_{\text{I}}$ and $\tilde{\Theta}_{\text{I}}$ are at most singletons.

However, the intermediary result requires more work when $\bar{\boldsymbol{\theta}} \in \bar{\Theta}_{\mathrm{I}}$ and $\tilde{\boldsymbol{\theta}} \in \tilde{\Theta}_{\mathrm{I}}$, but $\bar{\boldsymbol{\theta}} \neq \bar{\boldsymbol{\theta}}^0$, $\tilde{\boldsymbol{\theta}} \neq \tilde{\boldsymbol{\theta}}^0$. This case would arise when the data can be rationalized by the single equilibrium assumption despite that there are multiple equilibria in the data generating processes. For both $\bar{\boldsymbol{\theta}}^0$ and $\tilde{\boldsymbol{\theta}}^0$, notice that one can write the observable choice probabilities $p(\mathbf{y}|\mathbf{x}_{\mathrm{E}})$ as functions of $\boldsymbol{\theta}^0 = [\phi^0, \Upsilon^0]'$, i.e. the parameters associated with the true data generating process, and the corresponding unobservable-specific choice probabilities $p^0(\mathbf{y}|\boldsymbol{\nu}) \equiv \bar{p}^0(\mathbf{y}|\boldsymbol{\nu}) = \tilde{p}^0(\mathbf{y}|\boldsymbol{\nu})$. In fact:

$$p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right) = p^0\left(\mathbf{y}|\boldsymbol{\nu}^0\right) + \phi^0\left[p^0\left(\mathbf{y}|\boldsymbol{\nu}^1\right) - p^0\left(\mathbf{y}|\boldsymbol{\nu}^0\right)\right] \tag{55}$$

$$p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) = p^0\left(\mathbf{y}|\boldsymbol{\nu}^0\right) + \left[\Upsilon^0 + \phi^0\right]\left[p^0\left(\mathbf{y}|\boldsymbol{\nu}^1\right) - p^0\left(\mathbf{y}|\boldsymbol{\nu}^0\right)\right]. \tag{56}$$

By using (55), (56) one can rewrite (21), (22) for an arbitrary $\boldsymbol{\theta}$ as:

$$p\left(\mathbf{y}|\boldsymbol{\nu}^0\right) = p^0\left(\mathbf{y}|\boldsymbol{\nu}^0\right) + \left[\phi^0 - \Upsilon^0\frac{\phi}{\Upsilon}\right]\left[p^0\left(\mathbf{y}|\boldsymbol{\nu}^1\right) - p^0\left(\mathbf{y}|\boldsymbol{\nu}^0\right)\right] \tag{57}$$

$$p\left(\mathbf{y}|\boldsymbol{\nu}^1\right) = p^0\left(\mathbf{y}|\boldsymbol{\nu}^0\right) + \left[\phi^0 + \Upsilon^0\frac{1-\phi}{\Upsilon}\right]\left[p^0\left(\mathbf{y}|\boldsymbol{\nu}^1\right) - p^0\left(\mathbf{y}|\boldsymbol{\nu}^0\right)\right]. \tag{58}$$

Define the following functions:

$$b_0 \equiv p^0\left(0,0|\boldsymbol{\nu}^0\right)p^0\left(1,1|\boldsymbol{\nu}^0\right) - p^0\left(1,0|\boldsymbol{\nu}^0\right)p^0\left(0,1|\boldsymbol{\nu}^0\right); \tag{59}$$

$$\begin{aligned} b_1 \equiv\, & p^0\left(0,0|\boldsymbol{\nu}^0\right)\left[p^0\left(1,1|\boldsymbol{\nu}^1\right) - p^0\left(1,1|\boldsymbol{\nu}^0\right)\right] + p^0\left(1,1|\boldsymbol{\nu}^0\right)\left[p^0\left(0,0|\boldsymbol{\nu}^1\right) - p^0\left(0,0|\boldsymbol{\nu}^0\right)\right] \\ & - p^0\left(1,0|\boldsymbol{\nu}^0\right)\left[p^0\left(0,1|\boldsymbol{\nu}^1\right) - p^0\left(0,1|\boldsymbol{\nu}^0\right)\right] - p^0\left(0,1|\boldsymbol{\nu}^0\right)\left[p^0\left(1,0|\boldsymbol{\nu}^1\right) - p^0\left(1,0|\boldsymbol{\nu}^1\right)\right]; \end{aligned} \tag{60}$$

$$\begin{aligned} b_2 \equiv\, & \left[p^0\left(0,0|\boldsymbol{\nu}^1\right) - p^0\left(0,0|\boldsymbol{\nu}^0\right)\right]\left[p^0\left(1,1|\boldsymbol{\nu}^1\right) - p^0\left(1,1|\boldsymbol{\nu}^0\right)\right] \\ & - \left[p^0\left(0,1|\boldsymbol{\nu}^1\right) - p^0\left(0,1|\boldsymbol{\nu}^0\right)\right]\left[p^0\left(1,0|\boldsymbol{\nu}^1\right) - p^0\left(1,0|\boldsymbol{\nu}^0\right)\right]. \end{aligned} \tag{61}$$

By evaluating condition (54) at $p(\mathbf{y}|\boldsymbol{\nu}^0)$ given in (57), one gets a quadratic equation in $h_0(\boldsymbol{\theta}) \equiv \phi^0 - \Upsilon^0\phi/\Upsilon$ such that:

$$b_2 h_0(\boldsymbol{\theta})^2 + b_1 h_0(\boldsymbol{\theta}) + b_0 = 0. \tag{62}$$

Similarly, evaluating (54) at $p(\mathbf{y}|\boldsymbol{\nu}^1)$ given in (58), one obtains the following quadratic equation in $h_1(\boldsymbol{\theta}) \equiv \phi^0 + \Upsilon^0(1-\phi)/\Upsilon$:

$$b_2 h_1(\boldsymbol{\theta})^2 + b_1 h_1(\boldsymbol{\theta}) + b_0 = 0. \tag{63}$$

Since $\bar{\boldsymbol{\theta}} \in \bar{\Theta}_{\mathrm{I}}$ and $\tilde{\boldsymbol{\theta}} \in \tilde{\Theta}_{\mathrm{I}}$, (62) and (63) are both satisfied by $\bar{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$. Notice that the solutions to these equations will be functions of b_0, b_1 and b_2, which are the same for both $\bar{\boldsymbol{\theta}}$

and $\tilde{\boldsymbol{\theta}}$. There are at most two solutions for each quadratic equation. After fixing the sign of Υ, using an argument similar to what has been done above, one can show that there is at most one couple, say $h_0(\boldsymbol{\theta})^*$ and $h_1(\boldsymbol{\theta})^{**}$, that solves (62) and (63) simultaneously for a given $\boldsymbol{\theta}$. Therefore, as what is needed to show:

$$\bar{p}\left(\mathbf{y}|\boldsymbol{\nu}^0\right) = p^0\left(\mathbf{y}|\boldsymbol{\nu}^0\right) + h_0(\boldsymbol{\theta})^*\left[p^0\left(\mathbf{y}|\boldsymbol{\nu}^1\right) - p^0\left(\mathbf{y}|\boldsymbol{\nu}^0\right)\right] = \tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^0\right); \tag{64}$$

$$\bar{p}\left(\mathbf{y}|\boldsymbol{\nu}^1\right) = p^0\left(\mathbf{y}|\boldsymbol{\nu}^0\right) + h_1(\boldsymbol{\theta})^{**}\left[p^0\left(\mathbf{y}|\boldsymbol{\nu}^1\right) - p^0\left(\mathbf{y}|\boldsymbol{\nu}^0\right)\right] = \tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^1\right) \tag{65}$$

which holds $\forall \mathbf{y} \in \mathcal{Y}^2$.

I now turn to the proof of Corollary 2 *per se*. In particular, we want to show that, under the stated conditions, $\bar{\Theta}_{\mathrm{I}} \neq \emptyset$ implies $\tilde{\Theta}_I \neq \emptyset$. The converse is symmetric. Suppose that $\bar{\boldsymbol{\theta}} \in \bar{\Theta}_{\mathrm{I}}$ such that $\bar{\Theta}_{\mathrm{I}} \neq \emptyset$. To show that $\tilde{\Theta}_{\mathrm{I}} \neq \emptyset$, one must find a $\tilde{\boldsymbol{\theta}}$ that satisfies the conditions defining $\tilde{\Theta}_{\mathrm{I}}$. The converse of the intermediary result implies that the only candidate $\tilde{\boldsymbol{\theta}}$ must satisfy $\bar{p}(\mathbf{y}|\boldsymbol{\nu}) = \tilde{p}(\mathbf{y}|\boldsymbol{\nu})$ $\forall \mathbf{y} \in \mathcal{Y}^2$ and $\forall \boldsymbol{\nu} \in \mathcal{V}$. Such a $\tilde{\boldsymbol{\theta}}$ exists and is unique because: (a) (21) defines a one-to-one mapping between $\bar{p}(\mathbf{y}|\boldsymbol{\nu}^0)$ and $\tilde{\phi}/\tilde{\Upsilon}$; and (b) (22) defines a one-to-one mapping between $\bar{p}(\mathbf{y}|\boldsymbol{\nu}^1)$ and $\left(1-\tilde{\phi}\right)/\tilde{\Upsilon}$. Notice that since $\bar{\boldsymbol{\theta}} \in \bar{\Theta}_{\mathrm{I}}$, $\tilde{p}(\mathbf{y}|\boldsymbol{\nu}) = \bar{p}(\mathbf{y}|\boldsymbol{\nu})$ $\forall \mathbf{y} \in \mathcal{Y}^2$ and $\forall \boldsymbol{\nu} \in \mathcal{V}$ implies that $0 \leq \tilde{p}(\mathbf{y}|\boldsymbol{\nu}) \leq 1$ and rank $\left\{\tilde{\mathbf{P}}(\boldsymbol{\nu})\right\} = 1$ $\forall \boldsymbol{\nu} \in \mathcal{V}$. The only other restrictions that must be satisfied by $\tilde{\boldsymbol{\theta}}$ to belong to $\tilde{\Theta}_{\mathrm{I}}$ are $0 \leq \tilde{\Gamma}(\mathbf{x}_{\mathrm{E}}) \leq 1$ $\forall \mathbf{x}_{\mathrm{E}} \in \mathcal{X}_{\mathrm{E}}$. Again, without loss of generality, consider $\tilde{\Upsilon} > 0$.[12] Given $\mathbf{x}_{\mathrm{E}}$, these conditions can be written as:

$$\frac{-\tilde{\phi}}{\tilde{\Upsilon}} < \frac{\tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}) - \tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}{\tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - \tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}; \tag{66}$$

$$\frac{\tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}) - \tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}{\tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - \tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)} < \frac{1-\tilde{\phi}}{\tilde{\Upsilon}}. \tag{67}$$

From (21), one gets:

$$\frac{-\tilde{\phi}}{\tilde{\Upsilon}} = \frac{\tilde{p}(\mathbf{y}|\boldsymbol{\nu}^0) - \tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}{\tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - \tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}. \tag{68}$$

Similarly, from (22):

$$\frac{1-\tilde{\phi}}{\tilde{\Upsilon}} = \frac{\tilde{p}(\mathbf{y}|\boldsymbol{\nu}^1) - \tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}{\tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - \tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}. \tag{69}$$

It follows that (66) and (67) hold if and only if:

$$\frac{\tilde{p}(\mathbf{y}|\boldsymbol{\nu}^0) - \tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}})}{\tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - \tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)} < 0; \tag{70}$$

[12] Actually, since $\tilde{p}(\mathbf{y}|\boldsymbol{\nu}) = \bar{p}(\mathbf{y}|\boldsymbol{\nu})$, equation (75) below implies that:

$$\frac{\tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - \tilde{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}{\tilde{\Upsilon}} = \frac{\bar{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1) - \bar{p}(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0)}{\bar{\Upsilon}}.$$

Using both $\tilde{\Upsilon} > 0$ and $\bar{\Upsilon} > 0$ ensures that $\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) - \tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right)$ and $\bar{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) - \bar{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right)$ have the same signs.

$$\frac{\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{1}\right)-\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}\right)}{\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^{1}\right)-\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^{0}\right)}>0. \tag{71}$$

Notice that, by rearranging

$$\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}\right)=\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{0}\right)\left[1-\tilde{\Gamma}\left(\mathbf{x}_{\mathrm{E}}\right)\right]+\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{1}\right)\tilde{\Gamma}\left(\mathbf{x}_{\mathrm{E}}\right) \tag{72}$$

one can write:

$$\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{0}\right)-\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}\right)=-\left[\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{1}\right)-\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{0}\right)\right]\tilde{\Gamma}\left(\mathbf{x}_{\mathrm{E}}\right); \tag{73}$$

$$\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{1}\right)-\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}\right)=\left[\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{1}\right)-\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{0}\right)\right]\left[1-\tilde{\Gamma}\left(\mathbf{x}_{\mathrm{E}}\right)\right]. \tag{74}$$

Moreover, by evaluating (72) at $\mathbf{x}_{\mathrm{E}}^{0}$, $\mathbf{x}_{\mathrm{E}}^{1}$ and rearranging:

$$\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{1}\right)-\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{0}\right)=\frac{\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^{1}\right)-\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^{0}\right)}{\tilde{\Upsilon}}. \tag{75}$$

Therefore, $\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{1}\right)-\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{0}\right)$ and $\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^{1}\right)-\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^{0}\right)$ have the same sign. Suppose that $\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^{1}\right)-\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^{0}\right)>0$. From (75), $\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{1}\right)-\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{0}\right)>0$, (73) implies that $\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{0}\right)-\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}\right)<0$ and (74) implies that $\tilde{p}\left(\mathbf{y}|\boldsymbol{\nu}^{1}\right)-\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}\right)>0$. As a result, inequalities (70) and (71) hold. A similar argument can be used to show that these inequalities also hold when $\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^{1}\right)-\tilde{p}\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^{0}\right)<0$. It follows that $\tilde{\boldsymbol{\theta}}\in\tilde{\Theta}_{\mathrm{I}}$ and $\tilde{\Theta}_{\mathrm{I}}\neq\emptyset$, which completes the proof.

A.4 Proof of Proposition 2

Shi and Shum (2015, Theorem 4.1, p. 503) show that:

$$\liminf_{M\to\infty}\inf_{\boldsymbol{\theta}\in\Theta_{\mathrm{I}}}\Pr\left(\boldsymbol{\theta}\in\mathrm{CS}_{M}\left(\alpha\right)\right)\geq1-\alpha \tag{76}$$

where $\mathrm{CS}_{M}\left(\alpha\right)$ is defined as in (39). Under the null hypothesis that $\Theta_{\mathrm{I}}\neq\emptyset$, $\mathrm{E}_{\boldsymbol{\theta}}\left[\xi_{M}\right]=\Pr\left(\mathrm{CS}_{M}\left(\alpha\right)=\emptyset\right)$ for some $\boldsymbol{\theta}\in\Theta_{\mathrm{I}}$. Moreover, since $\mathrm{CS}_{M}\left(\alpha\right)=\emptyset\Rightarrow\boldsymbol{\theta}\notin\mathrm{CS}_{M}\left(\alpha\right)$ for any $\boldsymbol{\theta}\in\Theta_{\mathrm{I}}$, it follows that $\Pr\left(\boldsymbol{\theta}\notin\mathrm{CS}_{M}\left(\alpha\right)\right)\geq\Pr\left(\mathrm{CS}_{M}\left(\alpha\right)=\emptyset\right)$. Therefore:

$$\limsup_{M\to\infty}\sup_{\boldsymbol{\theta}\in\Theta_{\mathrm{I}}}\mathrm{E}_{\boldsymbol{\theta}}\left[\xi_{M}\right]\leq\limsup_{M\to\infty}\sup_{\boldsymbol{\theta}\in\Theta_{\mathrm{I}}}\Pr\left(\boldsymbol{\theta}\notin\mathrm{CS}_{M}\left(\alpha\right)\right) \tag{77}$$

$$=1-\liminf_{M\to\infty}\inf_{\boldsymbol{\theta}\in\Theta_{\mathrm{I}}}\Pr\left(\boldsymbol{\theta}\in\mathrm{CS}_{M}\left(\alpha\right)\right) \tag{78}$$

$$\leq\alpha \tag{79}$$

which completes the proof.

Appendix B Equalities and inequalities used for inference

In the simple $2 \times 2 \times 2$ case, there are two equalities that must be satisfied for $\boldsymbol{\theta}$ to belong to the identified set. These are:

$$a_2 \left[\frac{-\phi}{\Upsilon}\right]^2 - a_1 \frac{\phi}{\Upsilon} + a_0 = 0; \tag{80}$$

$$a_2 \left[\frac{1-\phi}{\Upsilon}\right]^2 + a_1 \frac{1-\phi}{\Upsilon} + a_0 = 0. \tag{81}$$

The remaining conditions defining the identified set are a list of inequalities ensuring that the unobserved probabilities involved in the finite mixture over $\boldsymbol{\nu}^0$ and $\boldsymbol{\nu}^1$ are all in the unit interval. These inequlities are:

$$0 < p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right) - \frac{\phi}{\Upsilon}\left[p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right)\right] < 1, \ \forall \mathbf{y} \in \mathcal{Y}^2; \tag{82}$$

$$0 < p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right) + \frac{1-\phi}{\Upsilon}\left[p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right)\right] < 1, \ \forall \mathbf{y} \in \mathcal{Y}^2; \tag{83}$$

$$0 < \phi + \Upsilon \frac{p\left(0,0|\mathbf{x}_{\mathrm{E}}\right) - p\left(0,0|\mathbf{x}_{\mathrm{E}}^0\right)}{p\left(0,0|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(0,0|\mathbf{x}_{\mathrm{E}}^0\right)} < 1, \ \forall \mathbf{x}_{\mathrm{E}} \in \mathcal{X}_{\mathrm{E}}. \tag{84}$$

One condition for the two-step method developped by Shi and Shum (2015) to be applicable is that the inequalities defining the identified set cannot depend on $\mathbf{p}$. Following their advice, I introduce slackness parameters that will transform inequalities depending on $\mathbf{p}$ into equalities. To do this, define $\eta_0\left(\mathbf{y}\right)$, $\eta_1\left(\mathbf{y}\right)$ $\forall \mathbf{y} \in \mathcal{Y}^2$ and $\eta\left(\mathbf{x}_{\mathrm{E}}\right)$ $\forall \mathbf{x}_{\mathrm{E}} \in \mathcal{X}_{\mathrm{E}}$ such that:

$$p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right) - \frac{\phi}{\Upsilon}\left[p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right)\right] - \eta_0\left(\mathbf{y}\right) = 0, \ \forall \mathbf{y} \in \mathcal{Y}^2; \tag{85}$$

$$p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right) + \frac{1-\phi}{\Upsilon}\left[p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right)\right] - \eta_1\left(\mathbf{y}\right) = 0, \ \forall \mathbf{y} \in \mathcal{Y}^2; \tag{86}$$

$$\phi + \Upsilon \frac{p\left(0,0|\mathbf{x}_{\mathrm{E}}\right) - p\left(0,0|\mathbf{x}_{\mathrm{E}}^0\right)}{p\left(0,0|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(0,0|\mathbf{x}_{\mathrm{E}}^0\right)} - \eta\left(\mathbf{x}_{\mathrm{E}}\right) = 0, \ \forall \mathbf{x}_{\mathrm{E}} \in \mathcal{X}_{\mathrm{E}}; \tag{87}$$

$$\eta_j\left(\mathbf{y}\right) > 0, \ 1 - \eta_j\left(\mathbf{y}\right) > 0, \ \forall \mathbf{y} \in \mathcal{Y}^2, \ j \in \{0,1\}; \tag{88}$$

$$\eta\left(\mathbf{x}_{\mathrm{E}}\right) > 0, \ 1 - \eta\left(\mathbf{x}_{\mathrm{E}}\right) > 0, \ \forall \mathbf{x}_{\mathrm{E}} \in \mathcal{X}_{\mathrm{E}}. \tag{89}$$

However, if one uses the equalities (80)-(81) and (85)-(87), the resulting Jacobian (with respect to $\mathbf{p}$) is not full row rank. To address this issue, I follow Shi and Shum (2015)'s recommendation. I drop the problematic equality constraints and rewrite them as inequality constraints on the slackness parameters. First notice that, for $j \in \{0,1\}$, $\sum_{\mathbf{y} \in \mathcal{Y}^2} p\left(\mathbf{y}|\,\boldsymbol{\nu}^j\right) = 1$ implies that:

$$\eta_j(0,0) + \eta_j(0,1) + \eta_j(1,0) + \eta_j(1,1) = 1. \tag{90}$$

Moreover, also for $j \in \{0,1\}$, (54) can be written as:

$$\eta_j(0,0)\,\eta_j(1,1) - \eta_j(0,1)\,\eta_j(1,0) = 0. \tag{91}$$

Therefore, for $j \in \{0,1\}$, one can use equations (90) and (91) to rewrite say $\eta_j(1,1)$ and $\eta_j(0,1)$ as functions of $\eta_j(0,0)$ and $\eta_j(1,0)$. In fact, it is easy to show that:

$$\eta_j(1,1) = \frac{\eta_j(1,0)}{\eta_j(0,0) + \eta_j(1,0)}\left[1 - \eta_j(0,0) - \eta_j(1,0)\right]; \tag{92}$$

$$\eta_j(0,1) = \frac{\eta_j(0,0)}{\eta_j(0,0) + \eta_j(1,0)}\left[1 - \eta_j(0,0) - \eta_j(1,0)\right]. \tag{93}$$

Given these expressions for $\eta_j(1,1)$, $\eta_j(0,1)$ and since $0 < \eta_j(0,0) < 1$ as well as $0 < \eta_j(1,0) < 1$, the condition:

$$1 - \eta_j(0,0) - \eta_j(1,0) > 0 \tag{94}$$

is sufficient to guarantee $\eta_j(1,1) > 0$ and $\eta_j(0,1) > 0$. Furthermore, $\eta_j(1,1) < 1$ and $\eta_j(0,1) < 1$ are guaranteed from the fact that:

$$\eta_j(0,0)\left[1 + \eta_j(1,0)\right] + \eta_j(1,0)^2 > 0; \tag{95}$$

$$\eta_j(1,0)\left[1 + \eta_j(0,0)\right] + \eta_j(0,0)^2 > 0 \tag{96}$$

always hold. Finally, notice that when evaluated at $\mathbf{x}_{\mathrm{E}}^0$ or $\mathbf{x}_{\mathrm{E}}^1$, (84) does not depend on $\mathbf{p}$. In fact, these inequalities are respectively equivalent to $0 < \phi < 1$ and $0 < \phi + \Upsilon < 1$.

To sum up, $\mathbf{g}^{\mathrm{e}}(\mathbf{p}, \boldsymbol{\theta})$ consists in:

$$a_2\left[\frac{-\phi}{\Upsilon}\right]^2 - a_1\frac{\phi}{\Upsilon} + a_0; \tag{97}$$

$$a_2\left[\frac{1-\phi}{\Upsilon}\right]^2 + a_1\frac{1-\phi}{\Upsilon} + a_0. \tag{98}$$

The vector $\mathbf{g}^{\mathrm{ie},1}(\mathbf{p}, \boldsymbol{\theta}) - \boldsymbol{\eta}$ contains:

$$p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right) - \frac{\phi}{\Upsilon}\left[p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right)\right] - \eta_0(\mathbf{y}),\ \mathbf{y} \in \{(0,0),(1,0)\}; \tag{99}$$

$$p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right) + \frac{1-\phi}{\Upsilon}\left[p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(\mathbf{y}|\mathbf{x}_{\mathrm{E}}^0\right)\right] - \eta_1(\mathbf{y}),\ \mathbf{y} \in \{(0,0),(1,0)\}; \tag{100}$$

$$\phi + \Upsilon\frac{p(0,0|\mathbf{x}_{\mathrm{E}}) - p\left(0,0|\mathbf{x}_{\mathrm{E}}^0\right)}{p\left(0,0|\mathbf{x}_{\mathrm{E}}^1\right) - p\left(0,0|\mathbf{x}_{\mathrm{E}}^0\right)} - \eta(\mathbf{x}_{\mathrm{E}}),\ \forall \mathbf{x}_{\mathrm{E}} \in \mathcal{X}_{\mathrm{E}} \setminus \left\{\mathbf{x}_{\mathrm{E}}^0, \mathbf{x}_{\mathrm{E}}^1\right\}. \tag{101}$$

The inequality conditions in $\mathbf{g}^{\text{ie},2}(\boldsymbol{\theta}, \boldsymbol{\eta})$ are:

$$\eta_j(\mathbf{y}),\ 1-\eta_j(\mathbf{y}),\ \mathbf{y} \in \{(0,0),(1,0)\},\ j \in \{0,1\}; \tag{102}$$

$$1-\eta_j(0,0)-\eta_j(1,0),\ j \in \{0,1\}; \tag{103}$$

$$\eta(\mathbf{x}_{\mathrm{E}}),\ 1-\eta(\mathbf{x}_{\mathrm{E}}),\ \forall \mathbf{x}_{\mathrm{E}} \in \mathcal{X}_{\mathrm{E}} \setminus \left\{\mathbf{x}_{\mathrm{E}}^{0}, \mathbf{x}_{\mathrm{E}}^{1}\right\}; \tag{104}$$

$$\phi; \quad 1-\phi; \quad \phi+\Upsilon; \quad \text{and} \quad 1-\phi-\Upsilon. \tag{105}$$

References

Aguirregabiria, V., and P. Mira (2007): "Sequential estimation of dynamic discrete games," *Econometrica*, 75(1), 1–53.

——— (2018): "Identification of games of incomplete information with multiple equilibria and common unobserved heterogeneity," *Working paper*.

Andrews, D., and P. Guggenberger (2009): "Validity of subsampling and "plug-in asymptotic" inference for parameters defined by moment inequalities," *Econometric Theory*, 25(3), 669–709.

Andrews, D., and G. Soares (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78(1), 119–157.

Aradillas-López, A. (2012): "Pairwise-difference estimation of incomplete information games," *Journal of Econometrics*, 168(1), 120–140.

Aradillas-López, A., and A. Gandhi (2016): "Estimation of games with ordered actions: An application to chain-store entry," *Quantitative Economics*, 7(3), 727–780.

Bajari, P., C. Benkard, and J. Levin (2007): "Estimating dynamic models of imperfect competition," *Econometrica*, 75(5), 1331–1370.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010): "Estimating static models of strategic interactions," *Journal of Business and Economic Statistics*, 28(4), 469–482.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp identification regions in models with convex moment predictions," *Econometrica*, 79(6), 1785–1821.

Bergemann, D., and S. Morris (2016): "Bayes correlated equilibrium and the comparison of information structures in games," *Theoretical Economics*, 11(2), 487–522.

Bonhomme, S., K. Jochmans, and J. Robin (2016): "Estimating multivariate latent-structure models," *The Annals of Statistics*, 44(2), 540–563.

Bugni, F., I. Canay, and X. Shi (2015): "Specification tests for partially identified models defined by moment inequalities," *Journal of Econometrics*, 185(1), 259–282.

Ciliberto, F., and E. Tamer (2009): "Market structure and multiple equilibria in airline markets," *Econometrica*, 6(77), 1791–1828.

De Paula, A., and X. Tang (2012): "Inference of signs of interaction effects in simultaneous games with incomplete information," *Econometrica*, 80(1), 143–172.

Galichon, A., and M. Henry (2011): "Set identification in models with multiple equilibria," *Review of Economic Studies*, 78(4), 1264–1298.

Ghanem, D. (2017): "Testing identifying assumptions in nonseparable panel data models," *Journal of Econometrics*, 197(2), 202–217.

Grieco, P. (2014): "Discrete games with flexible information structures: An application to local grocery markets," *RAND Journal of Economics*, 45(2), 303–340.

Hahn, J., H. Moon, and C. Snider (2017): "LM test of neglected correlated random effects and its application," *Journal of Business and Economic Statistics*, 35(3), 359–370.

Hall, P., and X. Zhou (2003): "Nonparametric estimation of component distributions in a multivariate mixture," *The Annals of Statistics*, 31(1), 201–224.

Henry, M., Y. Kitamura, and B. Salanié (2014): "Partial identification of finite mixtures in econometric models," *Quantitative Economics*, 5(1), 123–144.

Hsu, Y., C. Liu, and X. Shi (2018): "Testing generalized regression monotonicity," *Working paper*.

Hu, Y. (2008): "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, 144(1), 27–61.

——— (2017): "The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics," *Journal of Econometrics*, 200(2), 154–168.

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

Kasahara, H., and K. Shimotsu (2009): "Nonparametric identification of finite mixture models of dynamic discrete choices," *Econometrica*, 77(1), 135–175.

——— (2014): "Non-parametric identification and estimation of the number of components in multivariate mixtures," *Journal of the Royal Statistical Society - Series B*, 76(1), 97–111.

Kasy, M. (2015): "Nonparametric inference on the number of equilibria," *The Econometrics Journal*, 18(1), 1–39.

Khan, S., and D. Nekipelov (2018): "Information structure and statistical information in discrete response models," *Quantitative Economics*, 9(2), 995–1017.

Kitagawa, T. (2015): "A test for instrument validity," *Econometrica*, 83(5), 2043–2063.

Luo, Y., P. Xiao, and R. Xiao (2018): "Identification of dynamic games with unobserved heterogeneity and multiple equilibria: Global fast food chains in China," *Working paper*.

Magnolfi, L., and C. Roncoroni (2017): "Estimation of discrete games with weak assumptions on information," *Working paper*.

Milgrom, P., and R. Weber (1985): "Distributional strategies for games with incomplete information," *Mathematics of Operations Research*, 10(4), 619–632.

Mourifié, I., M. Henry, and R. Méango (2018): "Sharp bounds and testability of a Roy model of STEM major choices," *Working paper*.

Mourifié, I., and Y. Wan (2017): "Testing local average treatment effect assumptions," *Review of Economics and Statistics*, 99(2), 305–313.

Navarro, S., and Y. Takahashi (2012): "A semiparametric test of agent's information sets for games of incomplete information," *Working paper*.

Otsu, T., M. Pesendorfer, and Y. Takahashi (2016): "Pooling data across markets in dynamic Markov games," *Quantitative Economics*, 7(2), 523–559.

Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," *RAND Journal of Economics*, 38(2), 373–399.

Pesendorfer, M., and P. Schmidt-Dengler (2008): "Asymptotic least squares estimators for dynamic games," *Review of Economic Studies*, 75(3), 901–928.

Shi, X., and M. Shum (2015): "Simple two-stage inference for a class of partially identified models," *Econometric Theory*, 31(3), 493–520.

Sweeting, A. (2009): "The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria," *RAND Journal of Economics*, 40(4), 710–742.

Tamer, E. (2003): "Incomplete simultaneous discrete response model with multiple equilibria," *Review of Economic Studies*, 70(1), 147–165.

Xiao, R. (2018): "Identification and estimation of incomplete information games with multiple equilibria," *Journal of Econometrics*, 203(2), 328–343.