SHARP TEST FOR EQUILIBRIUM UNIQUENESS IN DISCRETE GAMES WITH PRIVATE INFORMATION AND COMMON KNOWLEDGE UNOBSERVED HETEROGENEITY

MATHIEU MARCOUX
Université de Montréal

This paper proposes a test of the single equilibrium in the data assumption commonly maintained when estimating static discrete games of incomplete information. By allowing for discrete common knowledge payoff-relevant unobserved heterogeneity, the test generalizes existing methods attributing all correlation between players' decisions to multiple equilibria. It does not require the estimation of payoffs and is therefore useful in empirical applications leveraging multiple equilibria to identify the model's primitives. 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. Under verifiable conditions, this testable implication is generically sufficient for degenerate equilibrium selection. The main identifying assumption is the existence of an observable variable that plays the role of a proxy for the unobservable heterogeneity. Examples of such proxies are provided based on empirical applications from the existing literature.

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. 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 a single equilibrium being realized in the data, an assumption also stated as the equilibrium selection mechanism being degenerate, is therefore desirable to guide applied researchers toward 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 both private information and common knowledge unobservables may be preferable in practice.

This paper provides a test of the single equilibrium in the data assumption, allowing for both private information and discrete common knowledge payoff-relevant unobservables (henceforth unobserved heterogeneity, for short).[1] Since the null hypothesis of interest depends on unobserved heterogeneity, it cannot be directly tested. The current paper derives sharp testable implications of this null hypothesis. The identification argument 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 partial identification results from Henry, Kitamura, and Salanié (2014) are used to derive sharp bounds for the distributions defining this finite mixture. Leveraging results from Kasahara and Shimotsu (2014), a degenerate equilibrium selection mechanism imposes further restrictions on these distributions. The identified set constructed from all these restrictions being nonempty is a testable implication of the null hypothesis of equilibrium uniqueness in the data generating process. This paper also provides conditions under which this testable implication is generically sufficient for degenerate equilibrium selection, i.e., the set of data generating processes for which the testable implication fails to be sufficient has Lebesgue measure zero. Moreover, the conditions defining the identified set are such that one can test the hypothesis of discrete unobserved heterogeneity separately from the equilibrium uniqueness assumption. 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 Chernozhukov, Lee, and Rosen's (2013) intersection bounds approach and simulation results suggest that it performs well.

---

[1] "Common knowledge unobservables" is often used in the literature on empirical games to describe information that is known to all players, but not to the econometrician. From a game theory point of view, these unobservables are "public information." In this paper, the former is preferred to avoid any confusion with existing work on empirical games.

How one treats multiple equilibria typically depends on how one is willing to interpret the information unobservable to the econometrician.[2] 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; Kline and Tamer, 2016; 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 the one hand, it is often assumed that the data have been generated by a single equilibrium at least for some state variables (e.g., Aguirregabiria and Mira, 2007; Bajari, Benkard, and Levin, 2007; Pakes, Ostrovsky, and Berry, 2007; Pesendorfer and Schmidt-Dengler, 2008; Bajari et al., 2010; Aradillas-López, 2012; Lewbel and Tang, 2015; etc.). 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 are therefore a valuable alternative to commonly used player-specific exclusion restrictions when the latter are not available in practice. To leverage this source of variation, one must provide evidence that there are multiple equilibria being realized in the data by rejecting the degenerate equilibrium selection mechanism assumption.

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 behavior 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 to solve for all the model's equilibria: the only relevant equilibrium is the one realized in the data and it 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.

---

[2]For recent reviews of the literature, see Berry and Tamer (2006), Bajari, Hong, and Nekipelov (2013), De Paula (2013), and Aradillas-López (2020).

Some tests of the single equilibrium in the data assumption have been proposed in the literature.[3] Two different approaches can be distinguished. The first one includes tests from 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 all unobservables from the econometrician's point of view are players' independent private information. However, 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 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 degenerate equilibrium selection hypothesis.

The second approach has the advantage of allowing for common knowledge unobserved heterogeneity. The test proposed here belongs to this category. Aguirregabiria and Mira (2019) and de Paula and Tang (2020) are the papers most closely related to the current one. A detailed discussion of each paper's respective contribution is given in Section 2. At this point, it is worth mentioning the main distinctions. As opposed to Aguirregabiria and Mira (2019), the test introduced below does not require estimating the payoff functions based on commonly used player-specific exclusion restrictions to separate multiple equilibria from common knowledge unobservable heterogeneity.[4] 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; de Paula and Tang, 2012; Aradillas-López and Gandhi, 2016). Of course, this advantage is not for 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 recent working paper by de Paula and Tang (2020) also avoids the need to estimate payoffs, but requires the researcher to group realizations of the game into clusters within which equilibrium selection is correlated if there are multiple equilibria. This grouping, typically based on some observables, is not needed in the test proposed here.[5]

The approach proposed in the current paper should be interpreted as a joint test of two assumptions: the single equilibrium in the data assumption and the

---

[3]The focus of the current paper is on static games. For dynamic games, Otsu, Pesendorfer, and Takahashi (2016) and Luo, Xiao, and Xiao (2022) have proposed tests which will be further discussed below.

[4]In that sense, the proposed test is aligned with a recommendation made by Sweeting (2009, p. 740): "[…] allowing for multiple equilibria can significantly increase computational costs […] and relying on multiple equilibria for identification may make the results even more dependent on the correct specification of the model. For these reasons, I see particular value in future research […] aimed at developing tests for the possible presence of multiple equilibria without requiring the full model to be estimated."

[5]More precisely, a grouping based on unobservables is built in the finite mixture identification results.

finite mixture representation of the unobserved heterogeneity. In particular, the test assumes the presence of common knowledge unobserved heterogeneity in the data generating process. An appealing by-product of the method is that the presence of unobserved heterogeneity can be separately tested without making assumptions on the number of equilibria realized in the data. I am not aware of the existence of such a test in the literature. In principle, one could therefore first test the relevance of unobserved heterogeneity to decide whether it should be taken into account when testing the single equilibrium in the data assumption in a second step. While this sequential approach would lead to possible pre-testing issues, providing a statistical method robust to these concerns is outside the scope of the current paper. In fact, this pre-testing problem would arise regardless whether one is using the method proposed here or another existing test of the single equilibrium in the data assumption in the second step. Alternatively, one may first use the joint test from the current paper and then, if this test rejects and one is worried that it may be due to the absence of unobserved heterogeneity, one could separately test the latter. Proceeding in this order does not de facto require a pre-testing correction in the case where the joint test is not rejected.

The rest of the paper is organized as follows. Related literature is summarized in Section 2, with 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. Monte Carlo simulations are reported in Section 6. Section 7 concludes. Proofs and further details are included in Appendixes.

2. RELATED LITERATURE

As mentioned above, a few papers from the literature on static games 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 have already been pointed out in Section 1. In fact, many of the existing tests (e.g., de Paula and Tang, 2012; Hahn et al. 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[6] (e.g., Hall and Zhou, 2003; Hu, 2008; Kasahara and Shimotsu, 2009, 2014;

---

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

Bonhomme, Jochmans, and Robin, 2016; and, the related but somewhat different identification arguments in Hu and Shum, 2012). 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. This lower bound corresponds to the number of components in the joint distribution of players' decisions represented as a finite mixture over the multiple equilibria. 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 based on conditional independence cannot be applied.[7] The main issue is that, if one finds the correlation between players' decisions to be nonzero or if one finds more than one component 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 parametric model that he estimates using data on grocery stores entry and exit. His results suggest that failing to include both private and common knowledge unobservable information may generate misleading results. More recently, Magnolfi and Roncoroni (2022) 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 very weak assumptions on the information structure of the game.[8] 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 and to allow for another source of possible correlation between players' decisions not due to multiple equilibria being realized in the data.

A few papers propose semi-parametric identification results that allow for multiple equilibria and common knowledge unobserved heterogeneity in static

---

[7]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.

[8]See Syrgkanis, Tamer, and Ziani (2018) for an econometric application of the same solution concept to auctions. While the information structure used in Magnolfi and Roncoroni (2022) is more flexible than the one considered in the current paper, point identification of the structural parameters under Bayes Correlated Equilibrium requires player-specific regressors which are not needed to perform the test proposed in the current paper. Moreover, the identification argument using multiple equilibria in the data leveraged in Sweeting (2009), de Paula and Tang (2012), and Aradillas-López and Gandhi (2016) are based on Bayesian Nash equilibria. The information structure considered here is more restrictive than Magnolfi and Roncoroni (2022), but still more flexible than games of pure incomplete information.

games of incomplete information. One of them is Khan and Nekipelov (2018) who show that the linear strategic interaction parameters are identified even if continuous unobserved heterogeneity is drawn from an unknown distribution. However, the identification results are obtained under the assumption that the equilibrium selection mechanism is uniform over the equilibria of the model and therefore does not vary with the common knowledge state variables. When testing the single equilibrium in the data assumption, it is preferable not to make such an assumption on the equilibrium selection mechanism. Moreover, their identification result requires observing large realizations of player-specific regressors. As mentioned above, testing the single equilibrium in the data assumption is especially desirable when such regressors are not available.

Another paper is Magesan (2018) who shows identification of the payoffs provided that the equilibrium selection mechanism is degenerate for some realizations of the game. In that sense, the test proposed here is complementary to his identification results.

As mentioned above, Aguirregabiria and Mira (2019) is closely related to the current paper. 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's (2019) main identification result may be problematic if one is solely interested in testing equilibrium uniqueness. There are two limitations worth pointing out. First, 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 its beliefs about its competitors' behavior (e.g., Bajari et al., 2010). However, one important motivation for testing equilibrium uniqueness is when such exclusion restrictions are not available and one would like to leverage multiplicity of equilibria as an alternative source of variation to identify payoffs. In such cases, one cannot apply Aguirregabiria and Mira's (2019) 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's (2019) sequential argument is given by the number of alternatives available in the

players' choice set, raised to the power $\left\lfloor \frac{N-1}{2} \right\rfloor$, 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 (2019) also propose nonsequential results which require the exclusion restrictions commonly used in empirical games to be sufficiently over-identifying. While their nonsequential approach has some advantages (e.g., it may allow for $N = 2$), it is still not applicable without variables satisfying these exclusion restrictions.

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 proposed here 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.

The second paper most closely related to the current one is the recent paper by de Paula and Tang (2020). They also propose testable implications for the single equilibrium in the data assumption which avoid having to estimate payoffs. This paper extends de Paula and Tang (2012) by allowing for players' private information to be correlated possibly via continuous unobserved heterogeneity. For static games, the test assumes that the researcher can group realizations of the game into clusters within which equilibrium selection is correlated if there are multiple equilibria. In other words, the arguments from de Paula and Tang (2012) apply within clusters which are defined by the researcher based on observables. To some extent, the finite mixture identification results leveraged in the test proposed below can be interpreted as a clustering which is based on unobservables that are left unspecified by the researcher (up to the support being finite and discrete) and other observable characteristics. Also, while the test proposed by de Paula and Tang (2020) is applicable to binary decisions, the test in the current paper allows for an arbitrary number of discrete choices.

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 subvectors 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), and Bonhomme et al. (2016). This conditional independence is also needed to

identify the number of mixtures when using Kasahara and Shimotsu's (2014) 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.[9] With this restriction, one can write the conditional component distributions and the mixing probability weights in terms of some set-identified parameters. Henry et al. (2014) introduced this approach and they also provided an extensive discussion suggesting that the required exclusion 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 work, 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 proposed here. Another related paper is Espin-Sanchez, Parra, and Wang (2022) who provide a testable condition that guarantees equilibrium uniqueness given estimates of the payoffs. Once again, estimating payoffs requires first taking a stance on whether a single or multiple equilibria are realized in the data.

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 et al., 2016, Sect. 3.5; Luo, Xiao, and Xiao, 2022; de Paula and Tang, 2020, Sect. 3). A nice feature of the dynamic setting is that the econometrician typically observes each market for more than one period, which gives some traction when trying to separate time-invariant unobserved heterogeneity from multiple equilibria. Unfortunately, this extra dimension, i.e., time, is often 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.

---

[9]Notice that this exclusion restriction does not correspond to the one commonly used to identify payoff functions as in Bajari et al. (2010). This is partly the reason why "proxy" will often be 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. A STATIC DISCRETE GAME WITH SIMULTANEOUS DECISIONS

This section describes the economic model, i.e., the game as it is played by the players, and its econometric counterpart.

3.1. Economic Model

Consider a game where N players, indexed by $i \in \{1, 2, \ldots, N\} = \mathcal{N}$, simultaneously choose from a finite and discrete choice set $\mathcal{Y} = \{0, 1, \ldots, J\}$. Players' decisions are stored in a vector of random variables $\mathcal{Y} = [\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_N]'$ with realizations $\mathbf{y} = [y_1, y_2, \ldots, y_N]' \in \mathcal{Y}^N$.[10]

Players' decisions are contingent on some state variables, which are separated into two categories depending on whether they are observed by all players. Let $\mathcal{S} = [\mathcal{S}_1', \mathcal{S}_2', \ldots, \mathcal{S}_N']'$ with realizations $\mathbf{s} = [\mathbf{s}_1', \mathbf{s}_2', \ldots, \mathbf{s}_N']' \in \mathscr{S}^N$ be some information that is common knowledge to all players. Furthermore, let $\mathcal{E} = [\mathcal{E}_1', \mathcal{E}_2', \ldots, \mathcal{E}_N']'$ with realizations $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', \ldots, \boldsymbol{\varepsilon}_N']'$ such that $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \ldots, \varepsilon_{iJ}]' \in \mathbb{R}^J$ be some private information. Let $G_{\mathcal{E}_i}(\cdot)$ denote the cumulative density function of $\mathcal{E}_i$. Because player i's opponents do not observe $\boldsymbol{\varepsilon}_i$, this is a game of incomplete information.

Let $\pi_i(\cdot) : \mathcal{Y}^N \times \mathscr{S} \times \mathbb{R} \mapsto \mathbb{R}$ be player i's payoff function. While the payoff of player i choosing $y_i = 0$ is normalized to 0, the payoff when choosing $y_i = j \neq 0$ is denoted by $\pi_i(j, \mathbf{y}_{-i}, \mathbf{s}_i, \varepsilon_{ij})$, where $\mathbf{y}_{-i}$ refers to the decisions of player i's opponents. The following assumption, which is common in the literature, is maintained on the payoff functions and the distributions of $\mathcal{S}, \mathcal{E}$.

> **Assumption 1** (State variables and payoffs).   (i) $\mathcal{S}, \mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_N$ are mutually independent.
> (ii) $G_{\mathcal{E}_i}(\cdot)$ $\forall i \in \mathcal{N}$ are common knowledge to all players and they are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^J$.
> (iii) $\pi_i(\cdot)$ $\forall i \in \mathcal{N}$ are common knowledge to all players.

It is worth mentioning that Assumption 1(i) is actually stronger than what will be required by the statistical test. Since the test will be performed after conditioning on a subvector of realizations of $\mathcal{S}$, $\mathcal{E}$ is only required to be independent of the remaining elements in $\mathcal{S}$. A more precise version of Assumption 1(i) is given in Assumption 2(iv).

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, all players simultaneously decide, i.e., $\mathbf{y}$ is realized and commonly observed. To sum up, at

---

[10]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.

the time of the simultaneous decisions, player i's information set is

$$\mathscr{I}_i = \left\{ \mathbf{s}, \boldsymbol{\varepsilon}_i, \pi_1(\cdot), \ldots, \pi_N(\cdot), G_{\boldsymbol{\mathcal{E}}_1}(\cdot), \ldots, G_{\boldsymbol{\mathcal{E}}_N}(\cdot) \right\}. \tag{1}$$

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

$$p_i(j|\mathbf{s}) = \int \mathbb{1}\{\sigma_i(\mathscr{I}_i) = j\} dG_{\boldsymbol{\mathcal{E}}_i}(\boldsymbol{\varepsilon}_i), \tag{2}$$

which can be interpreted as the beliefs of player i's opponents regarding player i's decision, when player i behaves according to strategy $\sigma_i(\mathscr{I}_i)$. Collect player i's probabilities in $\mathbf{p}_i(\mathbf{s}) = [p_i(1|\mathbf{s}), \ldots, p_i(J|\mathbf{s})]'$ and let $\mathbf{p}(\mathbf{s}) \equiv \left[\mathbf{p}_1(\mathbf{s})', \ldots, \mathbf{p}_N(\mathbf{s})'\right]'$. Let $\pi_i^{\mathbf{p}}(j, \mathbf{s}, \varepsilon_{ij})$ denote the expected payoff of player i choosing $y_i = j$, which is computed by integrating out $\mathbf{y}_{-i}$ using the corresponding elements of $\mathbf{p}(\mathbf{s})$. If each player's strategy is to maximize expected payoffs, (2) is equivalent to

$$p_i(j|\mathbf{s}) = \int \mathbb{1}\left\{\pi_i^{\mathbf{p}}(j, \mathbf{s}, \varepsilon_{ij}) - \pi_i^{\mathbf{p}}(k, \mathbf{s}, \varepsilon_{ik}) \geq 0, \forall k \in \mathscr{Y}\right\} dG_{\boldsymbol{\mathcal{E}}_i}(\boldsymbol{\varepsilon}_i). \tag{3}$$

The right-hand side of equation (3) is the best response mapping of player i given its beliefs regarding its opponents' decisions. Let $\psi_{ij}(\mathbf{s}, \cdot)$ denote this mapping, collect choice-specific mappings in $\boldsymbol{\psi}_i(\mathbf{s}, \cdot) = [\psi_{i1}(\mathbf{s}, \cdot), \psi_{i2}(\mathbf{s}, \cdot), \ldots, \psi_{iJ}(\mathbf{s}, \cdot)]'$ and define

$$\boldsymbol{\Psi}(\mathbf{s}, \cdot) \equiv \left[\boldsymbol{\psi}_1(\mathbf{s}, \cdot)', \boldsymbol{\psi}_2(\mathbf{s}, \cdot)', \ldots, \boldsymbol{\psi}_N(\mathbf{s}, \cdot)'\right]', \tag{4}$$

where $\boldsymbol{\Psi}(\mathbf{s}, \cdot) : [0, 1]^{JN} \mapsto [0, 1]^{JN}$ is a mapping in the probability space. Given $\boldsymbol{\Psi}(\mathbf{s}, \cdot)$, 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* $\mathbf{p}(\mathbf{s}) = \boldsymbol{\Psi}(\mathbf{s}, \mathbf{p}(\mathbf{s}))$.

Definition 1 simply states that, in equilibrium, players' beliefs are consistent with their opponents'. In fact, a BNE in the probability space is a fixed point of the best response mapping. Since $\boldsymbol{\Psi}(\mathbf{s}, \cdot)$ 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 $\mathbf{s} \in \mathscr{S}$. However, uniqueness is not guaranteed. Let $\mathscr{T}(\mathbf{s})$ be the set collecting all equilibria that the model admits at a given $\mathbf{s}$.

In order to fix ideas, it is worth introducing 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 operates in the market and $y_i = 0$ otherwise. In this case, $\boldsymbol{\mathcal{S}}$ could be some common knowledge
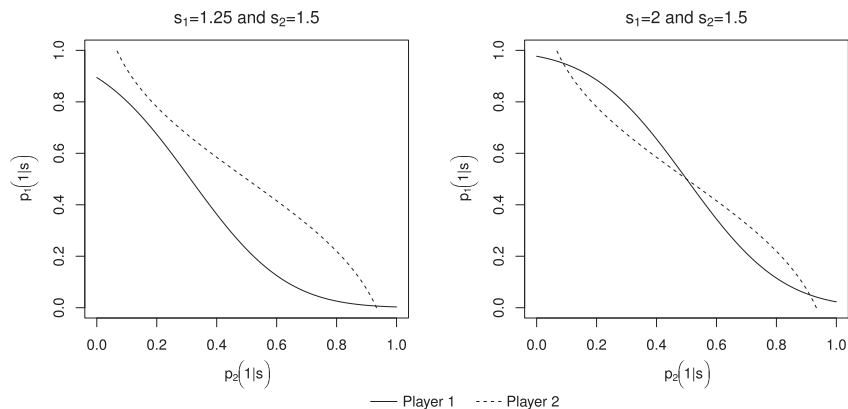
FIGURE 1. Equilibria multiplicity in Example 1.

information about market size and consumers' preferences. Moreover, $\mathcal{E}$ could refer to some private information cost shifters such as managerial ability. Let players' payoffs when operating in the market be

$$\pi_1\left(1, y_2, \mathbf{s}, \varepsilon_1\right) = s_1 - 4y_2 - \varepsilon_{11}, \tag{5}$$

$$\pi_2\left(1, y_1, \mathbf{s}, \varepsilon_2\right) = s_2 - 3y_1 - \varepsilon_{21}, \tag{6}$$

and player i's payoffs when $y_i = 0$ are normalized to 0. Furthermore, let $\boldsymbol{\varepsilon} = [\varepsilon_{11}, \varepsilon_{21}]$ be drawn from Normal $(\mathbf{0}, \mathbf{I}_2)$, 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

$$\boldsymbol{\Psi}\left(\mathbf{s}, \mathbf{p}\left(\mathbf{s}\right)\right) = \left[\begin{array}{c} \Phi\left(s_1 - 4p_2\left(1 | \mathbf{s}\right)\right) \\ \Phi\left(s_2 - 3p_1\left(1 | \mathbf{s}\right)\right) \end{array}\right]. \tag{7}$$

Figure 1 is the graphical representation of the best response mapping in (7) for all $\mathbf{p}\left(\mathbf{s}\right) \in [0, 1]^2$. The BNE(s) are given by the intersection(s) of the best response functions. This figure clearly illustrates that, for a given set of primitives, different realizations of $\mathcal{S}$ are associated with different BNE's and in particular with different numbers of BNE's.

3.2. Econometric Model

The game described so far is the *economic game* as it is played by the players. Consider now 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 $\mathcal{S}$. The following assumption is maintained on $\mathcal{S}$.

Assumption 2 (Common knowledge payoff-relevant state variables).   (i) $\mathcal{S} = [\mathcal{X}', \mathcal{V}']'$, where $\mathcal{X}$ is observable to the econometrician, but $\mathcal{V}$ is not.
(ii) $\mathcal{X}$ and $\mathcal{V}$ have finite and discrete supports $\mathscr{X} = \{\mathbf{x}^0, \mathbf{x}^1, \ldots, \mathbf{x}^{|\mathscr{X}|-1}\}$ and $\mathscr{V}(\mathbf{x}) = \{\boldsymbol{v}^0, \boldsymbol{v}^1, \ldots, \boldsymbol{v}^{|\mathscr{V}(\mathbf{x})|-1}\}$.
(iii) Realized $\boldsymbol{v}$'s are drawn from the conditional distribution with probability mass function $\gamma(\cdot | \mathbf{x})$.
(iv) Conditional on $\mathcal{X}$, $\mathcal{V}$ is independent from $\mathcal{E}$.

Typically, $\mathcal{V}$ can be thought of as a vector of common knowledge unobservable heterogeneity.[11] It is written as a vector-valued random variable to emphasize that it may capture different forms of unobserved heterogeneity. It could be a vector of player-specific unobservables or a scalar capturing unobservables that do not vary across players.

Finite and discrete unobserved heterogeneity is arguably a natural first step toward allowing for common knowledge unobserved heterogeneity when testing the single equilibrium in the data assumption. Currently available tests would reject the null hypothesis of a single equilibrium in the data if players' decisions were sufficiently correlated conditional on observed state variables. The test proposed here can be interpreted as a device to check whether sufficient correlation remains when allowing for *some* common knowledge unobserved heterogeneity. In that sense, allowing for a fixed number of possible realizations of unobservables (without constraining their distribution) is a crude, but useful approximation to detect correlation due to such unobservable information.

One appealing feature of the discrete unobserved heterogeneity in the current setting is that the support $\mathscr{V}(\mathbf{x})$ is allowed to vary with $\mathbf{x}$, such that different realizations of the game are allowed to have different realizations of discrete unobserved heterogeneity. Further, $|\mathscr{V}(\mathbf{x})|$ is identifiable from the data and the method allows the researcher to check whether the data reject assumptions maintained on the support of $\mathscr{V}(\mathbf{x})$, separately from the equilibrium uniqueness hypothesis. More details are provided in Section 4.

In order to appreciate the relevance of using discrete unobserved heterogeneity, it is worth discussing alternative approaches proposed in the literature to allow for common knowledge unobserved heterogeneity in static discrete games of incomplete information. First of all, several empirical applications ignore such unobserved heterogeneity by assuming that all unobservables from the point of view of the econometrician are private information. Of course, when the data has a panel structure in which realizations of the game are observed along more than one dimension (e.g., multiple geographical markets observed over multiple periods of time in a game of market entry), introducing fixed effects may capture some of the unobservable information known to all players. Unfortunately, this option is not possible if one observes a single cross section of replications of the game.

---

[11]Since the support of $\mathcal{V}$ depends on $\mathbf{x}$, it may be preferable to write $\mathcal{V}(\mathbf{x})$ and $\boldsymbol{v}(\mathbf{x})$. However, the argument is dropped to alleviate notation.

One alternative proposed in the literature has been to introduce some parametric common knowledge unobserved heterogeneity. For instance, in some of his specifications, Sweeting (2009) allows for player-specific unobserved heterogeneity constant across repetitions of the game which is drawn from a distribution known up to some parameters and enters players' payoffs additively. Another example is Bajari et al. (2010) who assume that common knowledge unobserved heterogeneity is simply a function of observable state variables.

The current approach does restrict the support of the common knowledge unobserved heterogeneity to be finite and discrete, which is not the case of the specifications proposed by Sweeting (2009) and Bajari et al. (2010). However, it has two considerable advantages. First, it does not require distributional assumptions beyond the support being finite and discrete. Second, the researcher does not have to specify how this common knowledge unobserved heterogeneity enters players' payoffs. In fact, $\mathcal{V}$ simply defines different components of a finite mixture representation of players' choice probabilities, without having to specify an expression for the distribution of players' decisions conditional on v.

It should be emphasized that finite mixture representations of common knowledge unobserved heterogeneity have been used when estimating dynamic discrete games. Both Aguirregabiria and Mira (2007) and Igami and Yang (2016) show that allowing for a finite number of points in the support of the common knowledge unobserved heterogeneity is able to capture confounding factors which, when ignored by the researcher, substantially bias strategic interaction estimates.

Notice that $\mathcal{X}$ being finite and discrete is also maintained by de Paula and Tang (2012) among others. While not essential in theory, this assumption is very convenient in practice since different realizations of $\mathcal{X}$ can be associated with different numbers of equilibria both in the model (see Figure 1) and in the data. This is the reason why the identification result holds for fixed realizations of $\mathcal{X}$. If $\mathcal{X}$ is continuous, one must take into account potential discontinuities in choice probabilities conditional on realizations of $\mathcal{X}$ that may be introduced by different numbers of equilibria across these realizations.[12]

Going back to the running example, $|\mathcal{V}(\mathbf{x})| = 2$ can be used to capture relevant unobservable information in a simple game of market entry.

Example 1 (Simple game of market entry, continued). In the simple game of market entry between two firms, $\mathcal{V}$ could capture some information related to unobservable demand and/or cost shifters. While both firms have potentially gathered information about these shifters, such information may remain unobservable to the econometrician or may be impossible to measure accurately. In this setting, even the simplest possible case allowing for two elements in the support of $\mathcal{V}$ helps to better interpret firms' decisions.

---

[12]As mentioned in Aguirregabiria and Mira (2019), nonparametric estimation of choice probabilities when the location of such discontinuities is unknown has been studied by Muller (1992) and Delgado and Hidalgo (2000).

Let $\mathscr{V}(\mathbf{x}) = \{\boldsymbol{v}^0, \boldsymbol{v}^1\}$. The interpretation of $\boldsymbol{v}^0$ and $\boldsymbol{v}^1$ depends on how they affect firms' probabilities of entering the market. For instance, $p_i\left(1|\mathbf{x}, \boldsymbol{v}^0\right) < p_i\left(1|\mathbf{x}, \boldsymbol{v}^1\right)$ for $i = 1, 2$ would suggest that if $\mathcal{V} = \boldsymbol{v}^1$, both firms are more likely to enter the market than if $\mathcal{V} = \boldsymbol{v}^0$. In other words, conditional on $\mathbf{x}$, $\mathcal{V}$ potentially reflects some market-level common knowledge unobserved heterogeneity. Alternatively, suppose that $p_1\left(1|\mathbf{x}, \boldsymbol{v}^0\right) < p_1\left(1|\mathbf{x}, \boldsymbol{v}^1\right)$ and $p_2\left(1|\mathbf{x}, \boldsymbol{v}^0\right) > p_2\left(1|\mathbf{x}, \boldsymbol{v}^1\right)$. One way to interpret $\mathcal{V}$ is that it captures some player-specific common knowledge unobserved heterogeneity that has a very different effect on firms' probabilities of entering the market given $\mathbf{x}$. While $\boldsymbol{v}^0$ decreases firm 1's probability of entry relative to $\boldsymbol{v}^1$, it increases firm 2's.

Finally, since the game may admit multiple equilibria, the econometric model requires an equilibrium selection mechanism assessing the probability that a given equilibrium is realized in the data. 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 are necessarily realized in the data. Let $\mathscr{T}^*(\mathbf{x}, \boldsymbol{v}) \subseteq \mathscr{T}(\mathbf{x}, \boldsymbol{v})$ be the subset of the model's equilibria that are realized in the data. Furthermore, let $\lambda(\cdot|\mathbf{x}, \boldsymbol{v})$ be the equilibrium selection mechanism given the payoff-relevant information observable to all players. Let $\mathcal{T}$ be a random variable labeling which equilibrium is played. More precisely, each equilibrium realized in the data is indexed by $\tau \in \mathscr{T}^*(\mathbf{x}, \boldsymbol{v})$.[13] The following assumption, which is also common in the literature (see for instance Aguirregabiria and Mira, 2019, Assum. 3, p. 1668) is maintained on $\mathcal{T}$. It ensures that players' private information does not affect which equilibrium is realized and, consequently, they cannot infer their competitors' private information from the realized equilibrium. It also rules out continuum of equilibria.

Assumption 3 (Equilibrium index). (*i*) $\mathcal{T}$ is independent from $\mathcal{E}_1, \ldots, \mathcal{E}_N$. (*ii*) $\mathcal{T}$ is discrete with support $\mathscr{T}^*(\mathbf{x}, \boldsymbol{v}) \equiv \{1, \ldots, |\mathscr{T}^*(\mathbf{x}, \boldsymbol{v})|\}$.

The framework described via Assumptions 1–3 actually allows for quite flexible correlations between variables that are unobservable from the econometrician's point of view. There are two different vectors of payoff-relevant unobservables, i.e., $\mathcal{E}$ and $\mathcal{V}$. Conditional on $\mathcal{X}$, common knowledge unobservable heterogeneity introduces correlation between players' payoff-relevant unobservables via $\mathcal{V}$ even if $\mathcal{E}_i$ are independent across players. Furthermore, $\mathcal{V}$ introduces correlation between payoff-relevant unobservables and the variable indicating which equilibrium is realized even if $\mathcal{T}$ is independent from $\mathcal{E}_1, \ldots, \mathcal{E}_N$.

Finally, the difference between $\mathscr{T}^*(\mathbf{x}, \boldsymbol{v})$ and $\mathscr{T}(\mathbf{x}, \boldsymbol{v})$ is key to interpret the degenerate equilibrium selection assumption. In fact, it corresponds to the special case $|\mathscr{T}^*(\mathbf{x}, \boldsymbol{v})| = 1$ for all $\boldsymbol{v} \in \mathscr{V}(\mathbf{x})$. In other words, for a given $\mathbf{x}$, $\lambda(\tau|\mathbf{x}, \boldsymbol{v}) = 1$ for the unique $\tau \in \mathscr{T}^*(\mathbf{x}, \boldsymbol{v})$ for all $\boldsymbol{v} \in \mathscr{V}(\mathbf{x})$. While the economic model may

---

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

admit multiple equilibria, this assumption requires the same equilibrium to be played whenever the same common knowledge payoff-relevant information is realized.

4. NONPARAMETRIC IDENTIFICATION

4.1. Data and Objects of Interest

When estimating an empirical game, the econometrician typically observes M realizations of the game. For each of these realizations, the data consist in $\{\mathbf{y}_m, \mathbf{x}_m : m = 1, \ldots, M\}$. There are therefore three random variables that are unobservable from the point of view of the researcher: (i) the private information shocks $\mathcal{E}$; (ii) the common knowledge variables $\mathcal{V}$; and (iii) the variable indicating which equilibrium is realized in the data $\mathcal{T}$. The assumptions maintained on the data generating process are formally stated in Assumption 4. Let $p(\mathbf{y}|\mathbf{x}, \boldsymbol{\nu}, \tau)$ denote the joint distribution of players' decisions conditional on $\mathbf{x}, \boldsymbol{\nu}, \tau$.

Assumption 4 (Data generating process). (*i*) $\{\boldsymbol{\varepsilon}_m, \mathbf{x}_m\}_{m=1}^{M}$ are independent and identically distributed across m. (*ii*) $\{\boldsymbol{\nu}_m\}_{m=1}^{M}$ are M independent draws from $\gamma(\cdot|\mathbf{x}_m)$. (*iii*) $\{\tau_m\}_{m=1}^{M}$ are M independent draws from $\lambda(\cdot|\mathbf{x}_m, \boldsymbol{\nu}_m)$. (*iv*) $\{\mathbf{y}_m\}_{m=1}^{M}$ are M independent draws from the multinomial distribution with vector of probabilities of success defined according to $p(\cdot|\mathbf{x}_m, \boldsymbol{\nu}_m, \tau_m)$.

Assumptions 1 to 4 imply that, conditional on $\{\mathbf{x}_m, \boldsymbol{\nu}_m, \tau_m\}_{m=1}^{M}$, y_{im} are independent across i and m. In fact, this conditional independence is all that is needed for a test of the single equilibrium in the data assumption based on the correlation between players' decisions after properly controlling for all players' common knowledge information.

Let $p(\mathbf{y}|\mathbf{x})$ be the conditional joint distribution of players' decisions that is point identified from the data. Given Assumptions 1 to 4, such distributions are double finite mixtures of the equilibrium conditional choice probabilities realized in the data. For a given $(\mathbf{y}, \mathbf{x}) \in \mathscr{Y}^N \times \mathscr{X}$:

$$p(\mathbf{y}|\mathbf{x}) = \sum_{\boldsymbol{\nu} \in \mathscr{V}(\mathbf{x})} p(\mathbf{y}|\mathbf{x}, \boldsymbol{\nu}) \gamma(\boldsymbol{\nu}|\mathbf{x}), \tag{8}$$

where

$$p(\mathbf{y}|\mathbf{x}, \boldsymbol{\nu}) = \sum_{\tau \in \mathscr{T}^*(\mathbf{x}, \boldsymbol{\nu})} p(\mathbf{y}|\mathbf{x}, \boldsymbol{\nu}, \tau) \lambda(\tau|\mathbf{x}, \boldsymbol{\nu}). \tag{9}$$

From equations (8) and (9), one can see that $p(\mathbf{y}|\mathbf{x})$ is a double finite mixture with a total of $\sum_{l=0}^{|\mathscr{V}(\mathbf{x})|-1} |\mathscr{T}^*(\mathbf{x}, \boldsymbol{\nu}^l)|$ components.

In the current setting, the main object of interest is the number of equilibria associated with each $\boldsymbol{\nu} \in \mathscr{V}(\mathbf{x})$ conditional on $\mathbf{x}$, i.e., each element of the set $\{|\mathscr{T}^*(\mathbf{x}, \boldsymbol{\nu}^l)|\}_{l=0}^{|\mathscr{V}(\mathbf{x})|-1}$. These objects of interest are used to test whether the single equilibrium in the data assumption holds at each realization of the finite and

discrete support of the common knowledge unobserved heterogeneity. Formally, the null and alternative hypotheses are stated as

$$\text{H}_0 : \left| \mathcal{T}^*(\mathbf{x}, \boldsymbol{\nu}) \right| = 1, \forall \boldsymbol{\nu} \in \mathcal{V}(\mathbf{x}); \quad \text{H}_1 : \left| \mathcal{T}^*(\mathbf{x}, \boldsymbol{\nu}) \right| > 1, \text{ for some } \boldsymbol{\nu} \in \mathcal{V}(\mathbf{x}).$$
$$\tag{10}$$

4.2. The Need for More Structure

The main challenge in testing the degenerate equilibrium selection mechanism while allowing for common knowledge unobserved heterogeneity is to separate each source of potential correlation in players' decisions, i.e., $\mathcal{T}$ and $\mathcal{V}$, conditional on $\mathcal{X}$. More precisely, it does not suffice to check the total number of components in the double finite mixture in (8) and (9), i.e., $\sum_{l=0}^{|\mathcal{V}(\mathbf{x})|-1} \left| \mathcal{T}^*(\mathbf{x}, \boldsymbol{\nu}^l) \right|$.[14] Testing H_0 rather requires being able to check the number of components *conditional* on the realizations of $\mathcal{V}$. The problem here is that H_0 depends on the unobservable heterogeneity and, therefore, cannot be tested directly.

To solve this problem, one must first put more structure on the double finite mixture of interest. The additional structure imposed in the proposed test is based on identification results from Henry et al. (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 one to separate the mixture over $\mathcal{V}$ from the mixture over $\mathcal{T}$. More importantly, it allows one to derive testable implications of H_0 without needing to estimate payoff functions.

4.3. 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 et al. (2014). The proxy variable interpretation will be highlighted later on. Once again, it is worth emphasizing that the following exclusion restriction is different from the one commonly used for the identification of discrete games.

Let the vector of observable state variables, $\mathcal{X}$, be divided into a subvector of variables that do not satisfy the exclusion restriction, $\mathcal{X}_{\text{NE}}$, and a subvector of variables that do satisfy it, $\mathcal{X}_{\text{E}}$, such that $\mathcal{X} = \left[\mathcal{X}_{\text{NE}}', \mathcal{X}_{\text{E}}' \right]'$, with realizations $\mathbf{x} = \left[\mathbf{x}_{\text{NE}}', \mathbf{x}_{\text{E}}' \right]' \in \mathcal{X}_{\text{NE}} \times \mathcal{X}_{\text{E}} \equiv \left\{ \mathbf{x}_{\text{NE}}^0, \mathbf{x}_{\text{NE}}^1, \ldots, \mathbf{x}_{\text{NE}}^{|\mathcal{X}_{\text{NE}}|-1} \right\} \times \left\{ \mathbf{x}_{\text{E}}^0, \mathbf{x}_{\text{E}}^1, \ldots, \mathbf{x}_{\text{E}}^{|\mathcal{X}_{\text{E}}|-1} \right\}$. Let $\tilde{\mathcal{X}}_{\text{E}}$ be a subset of $\mathcal{X}_{\text{E}}$ such that $\tilde{\mathcal{X}}_{\text{E}} \equiv \left\{ \tilde{\mathbf{x}}_{\text{E}}^0, \ldots, \tilde{\mathbf{x}}_{\text{E}}^{|\tilde{\mathcal{X}}_{\text{E}}|-1} \right\} \subseteq \mathcal{X}_{\text{E}}$ with $\tilde{\mathbf{x}}_{\text{E}}^0 \equiv \mathbf{x}_{\text{E}}^0$.

---

[14]For instance, under some conditions on $|\mathcal{V}|^N$, this total number of components could be identified using results from Kasahara and Shimotsu (2014). When one assumes away $\mathcal{V}$, then the total number of components in the "double" finite mixture becomes the number of equilibria realized in the data.

Similarly, let $\widetilde{\mathscr{Y}^N}$ be a subset of $\mathscr{Y}^N$ such that $\widetilde{\mathscr{Y}^N} \equiv \left\{ \tilde{\mathbf{y}}^1, \ldots, \tilde{\mathbf{y}}^{\left| \widetilde{\mathscr{Y}^N} \right|} \right\} \subseteq \mathscr{Y}^N$.
Assumption 5 formally states the identifying restrictions.

Assumption 5 (Identifying restrictions). For any $\mathbf{y} \in \mathscr{Y}^N$, $\mathbf{x}_{\mathrm{NE}} \in \mathscr{X}_{\mathrm{NE}}$, $\mathbf{x}_{\mathrm{E}} \in \mathscr{X}_{\mathrm{E}}$, and $\mathbf{v} \in \mathscr{V}(\mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}})$:

(i) (Support independence) $\mathscr{V}(\mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}})$ does not depend on $\mathbf{x}_{\mathrm{E}}$;
(ii) (Cardinality of the support) $|\mathscr{V}(\mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}})| \leq \min\left\{ (J+1)^N, |\mathscr{X}_{\mathrm{E}}| \right\}$;
(iii) (Relevance) $\gamma(\mathbf{v}|\mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}})$ depends on $\mathbf{x}_{\mathrm{E}}$;
(iv) (Redundancy) $p(\mathbf{y}|\mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}}, \mathbf{v})$ does not depend on $\mathbf{x}_{\mathrm{E}}$;
(v) (Sufficient variability in mixture weights) for an appropriate subset $\tilde{\mathscr{X}}_{\mathrm{E}}$, the square matrix with (i,j)th element given by $\gamma\left(\mathbf{v}^i \middle| \mathbf{x}_{\mathrm{NE}}, \tilde{\mathbf{x}}_{\mathrm{E}}^j \right) - \gamma\left(\mathbf{v}^i \middle| \mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}}^0 \right)$ for $i,j = 1, \ldots, |\mathscr{V}(\mathbf{x}_{\mathrm{NE}})| - 1$ is invertible; and
(vi) (Sufficient variability in unknown choice probabilities) for an appropriate subset $\widetilde{\mathscr{Y}^N}$, the square matrix with (i,j)th element given by $p\left(\tilde{\mathbf{y}}^i \middle| \mathbf{x}_{\mathrm{NE}}, \mathbf{v}^j \right) - p\left(\tilde{\mathbf{y}}^i \middle| \mathbf{x}_{\mathrm{NE}}, \mathbf{v}^0 \right)$ for $i,j = 1, \ldots, |\mathscr{V}(\mathbf{x}_{\mathrm{NE}})| - 1$ is invertible.

The support independence condition implies that the set of values of the mixing variables realized with a positive probability does not vary with $\mathscr{X}_{\mathrm{E}}$. This condition is important: in order to use variation in $\mathscr{X}_{\mathrm{E}}$ to identify the finite mixture over $\mathcal{V}$, such variation should not generate changes in the set of possible realizations of the mixing variable. Slightly abusing notation, $\mathscr{V}(\mathbf{x}_{\mathrm{NE}})$ 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 $\mathcal{Y}$ and in $\mathscr{X}_{\mathrm{E}}$ for the exclusion restriction approach to be able to capture all the relevant realizations of $\mathcal{V}$. The rationale for $\min\left\{ (J+1)^N, |\mathscr{X}_{\mathrm{E}}| \right\}$ will become clear in Lemma 1. In a game where two players choose between two actions, the test allows for up to $2^2 = 4$ different realizations of common knowledge unobserved heterogeneity, given $\mathbf{x}_{\mathrm{NE}}$, provided that there is enough variation in $\mathscr{X}_{\mathrm{E}}$. For a game where two players choose between three actions, there can be up to nine values of common knowledge unobserved heterogeneity. If three players choose between two actions, this number is 8. In other words, even for very simple games, the test allows for an appreciable number of realizations of common knowledge unobserved heterogeneity, which is exponentially growing in the number of players. As will be discussed in Section 4.4, applying the test to existing examples from the literature suggests that several unobservable realizations can be accommodated for when testing the single equilibrium in the data assumption.

The relevance condition requires the distribution of the unobservable $\mathcal{V}$ to depend on both $\mathscr{X}_{\mathrm{NE}}$ and $\mathscr{X}_{\mathrm{E}}$. Notice that this condition does not contradict the support independence one as long as realizations of $\mathcal{V}$ do not become zero probability events for some realization of $\mathscr{X}_{\mathrm{E}}$.

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 $\mathcal{X}_{\mathrm{E}}$ after conditioning on $\mathcal{X}_{\mathrm{NE}}$ and $\mathcal{V}$.[15] Once again, slightly abusing notation, such independence is made obvious by using $p(\mathbf{y}|\mathbf{x}_{\mathrm{NE}}, \boldsymbol{v})$, $p(\mathbf{y}|\mathbf{x}_{\mathrm{NE}}, \boldsymbol{v}, \tau)$, $\lambda(\tau|\mathbf{x}_{\mathrm{NE}}, \boldsymbol{v})$ and $\mathscr{T}^*(\mathbf{x}_{\mathrm{NE}}, \boldsymbol{v})$. In other words, $\mathcal{X}_{\mathrm{E}}$ provides some information about the distribution of the unobservable $\mathcal{V}$, but would not provide any information about players' decisions nor the equilibrium selection if $\mathcal{V}$ would be observable.

Finally, the assumption of sufficient variability in mixture weights and in unknown choice probabilities can be interpreted as rank conditions ensuring that the variation induced by $\mathcal{X}_{\mathrm{E}}$ and $\mathcal{Y}$ are sufficient to identify the mixture over $\mathcal{V}$.

4.4. Examples of Proxy Variables

Of course, a natural question to ask at this point is whether variables satisfying restrictions stated in Assumption 5 are easy to find. Remember that $\mathcal{V}$ 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. In particular, since one is not interested in quantifying the structural effects of such common knowledge unobserved heterogeneity per se, the test does not require precise measurements of these unobservables.

There are at least three categories of variables that can be used as proxies for common knowledge unobserved heterogeneity: imperfect measurements for suspected unobserved heterogeneity, number of players and predetermined outcomes. The following examples, which are based on existing empirical applications, highlight the potential for multiple equilibria, the relevance of discrete unobserved heterogeneity and the proxies that could be used.[16]

Depending on the empirical application, some of these three categories of variables may be more useful than others. For instance, if the researcher has a clear understanding of the type of common knowledge payoff-relevant unobserved heterogeneity that should be controlled for, imperfect measurements of this suspected unobserved heterogeneity may be quite appealing. As discussed below, there are

---

[15]Notice that the redundancy condition implies independence between $\mathcal{Y}$ and $\mathcal{X}_{\mathrm{E}}$ after conditioning on $\mathcal{X}_{\mathrm{NE}}$ and $\mathcal{V}$. 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, since the distribution of $\mathcal{V}$ depends on $\mathcal{X}_{\mathrm{E}}$, one cannot use the independence between $\mathcal{Y}$ and $\mathcal{X}_{\mathrm{E}}$ to identify the mixture over $\mathcal{V}$ as in the conditional independence approach.

[16]Henry et al. (2014, Sect. 1.2) discuss several examples of variables that can be used as proxies for unobservables in different empirical applications. Part of their discussion covers unobserved heterogeneity in structural microeconomic models. In their oligopoly pricing example, Henry et al. (2014, Appendix A) suggest using measures of realized profits, which are ex ante unknown to the players, to control for unobserved demand variables. Such proxies would indeed be informative about unobserved heterogeneity beyond $\mathcal{X}_{\mathrm{NE}}$ (relevance) as realized profits typically depend on common knowledge unobservables. However, it is not obvious that they can be applied in the current setting since realized profits would be informative about the equilibrium selection mechanism.

several examples of empirical applications where imperfect measurements are included as regressors to control for unobserved heterogeneity. If the nature of the unobserved heterogeneity is either not obvious or hard to control for using imperfect measurements, the number of players or predetermined outcomes may be more interesting options.

Imperfect Measurements of Suspected Unobserved Heterogeneity.   In empirical applications, researchers may suspect a specific source of unobserved (or even unmeasurable) heterogeneity to act as a confounding factor likely to bias the estimation. Including imperfect measurements of these unobservables is frequently done to control for such confounding factors at least partially. Commonly used examples from different fields of economics are test scores to control for cognitive abilities, income for well-being, value of houses for wealth, etc.

For instance, in a game of market entry, unobservable cost and demand shifters may make a market relatively more profitable than what is actually captured by observable payoff-relevant variables. When players' payoffs have a reduced form interpretation, i.e., when demand and supply are not explicitly modeled, one may therefore choose to include some imperfect measurements for these shifters.

If observable variables are suitable controls for the unobservables they are imperfectly capturing, they are likely to satisfy the conditions required for $\mathcal{X}_E$ to be valid excluded regressors. While they are imperfect measurements of the suspected unobserved heterogeneity, they are typically strongly correlated with it and therefore satisfy the relevance condition. However, conditional on these unobservables (and $\mathcal{X}_{NE}$), the imperfect measurements generally do not affect payoffs nor equilibrium selection. In other words, imperfect measurements should not be more informative about players' decisions nor the equilibrium being realized in the data than the actual unobservables (given $\mathcal{X}_{NE}$). It follows that the redundancy condition is satisfied as well.

It is important to note that the variable $\mathcal{X}_E$ is also allowed to be correlated with $\mathcal{X}_{NE}$ (just like $\mathcal{V}$ possibly is). However, for an imperfect measurement to be a valid proxy for $\mathcal{V}$, it must be such that it does not directly affect players' decisions conditional on $\mathcal{X}_{NE}$ and $\mathcal{V}$: the only reason why the imperfect measurement $\mathcal{X}_E$ is informative about players' decisions (conditional on $\mathcal{X}_{NE}$) is because $\mathcal{V}$ is not observed. Otherwise, this imperfect measurement should have been included in $\mathcal{X}_{NE}$.

Example 2 (Imperfect measurements of demand and cost shifters). Aradillas-López and Gandhi (2016) propose a novel estimator which uses correlation between players' decisions generated by multiple equilibria being realized in the data to estimate games of ordered actions. In their empirical application, they consider a game between the three main chains of retail pharmacies in the US: CVS ($i = 1$), Rite Aid ($i = 2$), and Walgreens ($i = 3$). Each player decides how many stores to operate in each geographical market (a core-based statistical area).

Aradillas-López and Gandhi (2016) justify the need to allow for multiple equilibria being realized in the data by the nature of the strategy space. They provide

simulation evidence that the relative large choice set of the players implies that the model admits multiple equilibria. Their estimation method effectively detects some correlation between players' decisions which they interpret as evidence of multiple equilibria being realized in the data.

Of course, correlation between players decisions could also be generated by common knowledge unobserved heterogeneity. However, Aradillas-López and Gandhi (2016, p. 755) argue that, in this industry, there is no "obvious, compelling demand side unobservable at the market level (e.g., an unexplained taste for health) that cannot be conditioned out with observables (such as the number of doctors in the market)." The observables included in $\mathcal{X}$ in their estimation are: population, average income per household, population density, median age in the population, total number of business establishments, and the distance to the nearest distribution center of each player. The authors note that the number of business establishments is included "to control for supply side unobservables" such as costs related to zoning restrictions.

The method proposed here could be used to formally test for multiple equilibria being realized in the data using the total number of business establishments as $\mathcal{X}_E$ to control for supply side unobservables such as unobserved zoning restrictions. While this imperfect measurement is correlated with the supply cost shifters suspected to introduce correlation between players' decisions, it is worth noting that it does not have a direct impact on pharmacies' payoffs nor the equilibrium being selected beyond the unobserved demand and cost shifters. For instance, it is not the total number of business establishments that makes a market more profitable, but rather zoning restrictions imperfectly measured by the number of business establishments, given all the other regressors included in $\mathcal{X}$.

This source of suspected unobserved heterogeneity is an example of market-level unobservable. Suppose that each firm chooses between 0, 1, or 2 stores. Since, $N = 3$, the test allows for up to $|\mathcal{V}(\mathbf{x}_{NE})| = 27$ different types of markets at a given realization of $\mathcal{X}_{NE}$, provided that there is enough variation in $\mathscr{X}_E$.

It is worth emphasizing that introducing some imperfect measurements to control for unobservables is very common when estimating games. For instance, in her analysis of location choices in the video retail industry, Seim (2006, pp. 629–630) uses "business density as a catch-all proxy for the general business environment." When studying competition in the airline industry, Ciliberto and Tamer (2009, p. 1808) "compute the sum of the geographical distances between a market's endpoints and the closest hub of a carrier as a proxy for the cost that a carrier has to face to serve that market," because "firm- and market-specific measures of cost are not available." In fact, in many empirical applications, geographical distance is often interpreted as some imperfect measurement of unobserved costs or scale economies. Another example is the distance between stores in Jia's (2008) analysis of competition between chain stores and discount retailers. Distance between stores is included in the payoff function because "nearby stores split the costs of operation, delivery, and advertising to achieve scale economies" (Jia, 2008,

p. 1276). In all these cases, the researcher has a clear interpretation of the unobserved heterogeneity to be controlled for.

An important remark is worth making at this stage. If proxy variables ($\mathcal{X}_E$) are already used to control for common knowledge unobserved heterogeneity ($\mathcal{V}$) when estimating empirical games, what prevents one from testing whether there is a single component in the mixture over realized equilibria directly from $p(\mathbf{y}|\mathbf{x}_{NE}, \mathbf{x}_E)$ instead of having to recover $p(\mathbf{y}|\mathbf{x}_{NE}, \mathbf{v})$? Presumably, one could work directly with $p(\mathbf{y}|\mathbf{x}_{NE}, \mathbf{x}_E)$ if there was a one-to-one relationship between $\mathbf{v}$ and $\mathbf{x}_E$, given $\mathbf{x}_{NE}$. While this one-to-one relationship may hold in some cases (e.g., using distance to proxy for transportation costs), it is far-fetched in others (e.g., total number of business establishments to control for supply side unobservables). When the relationship is not one-to-one, $p(\mathbf{y}|\mathbf{x}_{NE}, \mathbf{x}_E) \neq p(\mathbf{y}|\mathbf{x}_{NE}, \mathbf{v})$, but the latter probability is needed to test the single equilibrium in the data assumption. Of course, this inequality is not problematic when one is solely interested in estimating payoffs: controlling for confounding factors does not require a one-to-one relationship between $\mathbf{v}$ and $\mathbf{x}_E$, conditional on $\mathbf{x}_{NE}$.

When one does not have a clear a priori on the potential source of unobserved heterogeneity, one cannot find an imperfect measurement for such unobservables. For those cases, one could consider the following two other possible types of proxies that satisfy the conditions associated with $\mathcal{X}_E$.

Number of Players.   In some empirical applications, there may be variation in the number of players (e.g., potential entrants in a game of market entry) across different realizations of the game. Realizations of the game that are observationally equivalent in $\mathcal{X}_{NE}$, but are associated with different numbers of players, suggest that the researcher may fail to observe some information known to all players (relevance). A similar argument is often used to detect or to control for unobserved heterogeneity in auctions.[17] Under some conditions, the number of players does not affect the equilibrium of the game and is therefore not informative about which equilibrium will be realized beyond $\mathcal{X}_{NE}$ and $\mathcal{V}$ (redundancy). This would be the case for games between symmetric players with strategic interactions depending linearly on the fraction of competitors making the same decision.

Formally, let the payoff of player i choosing j be

$$\pi_i\left(j, \mathbf{y}_{-i}, \mathbf{x}_{NE}, \mathbf{v}, \varepsilon_{ij}\right) = \mu\left(j, \mathbf{x}_{NE}, \mathbf{v}, \varepsilon_{ij}\right) + \alpha \frac{\sum_{k \neq i} \mathbb{1}\left\{y_k = j\right\}}{N - 1}, \tag{11}$$

which is identical across players (up to the private information shock ε_{ij}) and depends linearly on the fraction of the $N-1$ competitors also choosing j, i.e., $\sum_{k \neq i} \mathbb{1}\{y_k = j\}/(N-1)$. Since players are symmetric, the probability of choosing j given $\mathbf{x}_{NE}, \mathbf{v}$, i.e., $p(j|\mathbf{x}_{NE}, \mathbf{v})$, does not vary across players. Player i's expected payoff of choosing j therefore becomes

---

[17]See, for example, Athey and Haile (2002), Hong and Shum (2002), Haile, Hong, and Shum (2003), Campo, Perrigne, and Vuong (2003), and Guerre, Perrigne, and Vuong (2009).

$$\pi_i^{\mathbf{P}}\left(j, \mathbf{x}_{\mathrm{NE}}, \boldsymbol{v}, \varepsilon_{ij}\right) = \mu\left(j, \mathbf{x}_{\mathrm{NE}}, \boldsymbol{v}, \varepsilon_{ij}\right) + \alpha \frac{(N-1)\, p\left(j \mid \mathbf{x}_{\mathrm{NE}}, \boldsymbol{v}\right)}{N-1} \tag{12}$$

$$= \mu\left(j, \mathbf{x}_{\mathrm{NE}}, \boldsymbol{v}, \varepsilon_{ij}\right) + \alpha\, p\left(j \mid \mathbf{x}_{\mathrm{NE}}, \boldsymbol{v}\right), \tag{13}$$

which does not depend on N. Since the best response mapping defining a BNE in the probability space (see Definition 1) is a function of differences of expected payoffs, the fact that expected payoffs do not depend on N implies that the equilibrium (equilibria) of the game is (are) necessarily independent of N.

The following example borrows from the model of commercial radio stations' advertisement timing decisions in Sweeting (2009). In particular, the case of symmetric stations is discussed in Sweeting (2009, p. 722) to highlight the identifying power of multiple equilibria realized in the data. The assumption of symmetric players may be imposed in empirical applications for which player-specific observables are either not available or do not induce significant variation in payoffs.

Example 3 (Varying numbers of symmetric players across realizations of the game). Consider a game in which N radio stations simultaneously choose between $J + 1$ possible time slots to air radio advertisements. The payoff of station i choosing j is

$$\pi_i\left(j, \mathbf{y}_{-i}, \mathbf{x}_{\mathrm{NE}}, v, \varepsilon_{ij}\right) = \mathbf{x}_{\mathrm{NE}}'\boldsymbol{\beta}_j + \alpha \frac{\sum_{k \neq i} \mathbb{1}\left\{y_k = j\right\}}{N-1} + v + \varepsilon_{ij}. \tag{14}$$

For instance, $\mathbf{x}_{\mathrm{NE}}$ could include observable demographics of the geographic market in which radio stations compete and v could represent some information known to all stations, but unobservable to the econometrician such as regional-specific listening habits. Since stations are identical up to their private information shock (ε_{ij}), $p\left(j \mid \mathbf{x}_{\mathrm{NE}}, v\right)$ does not vary with i. Using the same argument as above, expected payoffs and stations' best responses do not depend on N which implies that the symmetric BNE (equilibria) is (are) also independent of N.

The timing decision studied by Sweeting (2009) is in fact a coordination game. Radio stations have an incentive to air their commercial breaks simultaneously to discourage listeners from switching between different channels to avoid advertisements. The model therefore admits multiple equilibria and different equilibria may be realized in the data.

If the number of radio stations (N) varies across realizations of the game, it can be used as $\mathcal{X}_{\mathrm{E}}$. If two different regional markets with the same realization of $\mathcal{X}_{\mathrm{NE}}$ have different N's, it could be because they differ in $\mathcal{V}$ (relevance). Moreover, payoffs being symmetric across players and depending linearly on the fraction of competitors making the same decision imply that variation in N does not affect the equilibrium of the game (conditional on the realized $\mathcal{X}_{\mathrm{NE}}$, $\mathcal{V}$). It follows that, if one could observe the realized $\mathcal{V}$, N would only provide redundant information about stations' decisions and equilibrium selection (redundancy).

Since choice probabilities do not vary across players, the maximum number of points in the support of $\mathcal{V}$ is now given by $\min\{J+1, |\mathscr{X}_{\mathrm{E}}|\}$ instead of

$\min\left\{(J+1)^N, |\mathscr{X}_E|\right\}$. More details about this bound are available in the discussion following Lemma 1. It therefore linearly depends on the observed variation in N and the number of possible time slots $(J+1)$ considered in the analysis. In Sweeting (2009), $J = 2$ and, depending on market and station definitions, the average N varies between 4.9 and 13.3 stations per market. One could therefore identify up to three points of support in $\mathcal{V}$.

One advantage of using the number of players as a proxy for the unobserved heterogeneity is that, as opposed to imperfect measurements, it does not require knowing what is the source of heterogeneity one would like to control for. As long as the unobserved heterogeneity known to all players induces variation in the number of players across replications of the game that have the same realizations of $\mathcal{X}_{NE}$, the number of players is informative about $\mathcal{V}$ regardless of what $\mathcal{V}$ captures. As mentioned above, this is also the reason why variation in the number of bidders is often used to control for common knowledge unobserved heterogeneity in auctions.

It is also worth emphasizing that the requirement of players being symmetric, which is needed for the redundancy condition to hold, is not particularly restrictive in the current setting. In fact, testing the degenerate equilibrium selection assumption is relevant in applications where one would like to leverage variation in realized equilibria to identify strategic interactions between players. This source of variation is especially useful when players' payoffs are symmetric (up to their private information shocks) since, in those cases, player-specific variation in payoffs is not observed by the researcher. In other words, the need to test for the single equilibrium in the data assumption is often justified by the symmetry in players' payoffs. Moreover, beyond the example from Sweeting (2009) mentioned above, strategic interactions depending linearly on the fraction of competitors making the same decisions is very common in empirical applications of peer effects and social interactions since Brock and Durlauf (2001, p. 239) who refer to such a specification as "proportional spillovers."

Predetermined Outcomes.   Suppose that there are some common knowledge payoff-relevant variables which have the same effect on payoffs regardless of players' decisions given $\mathcal{X}_{NE}$ and $\mathcal{V}$. Under some conditions, such variables are valid candidates for proxies. More precisely, let $\mathcal{X}_E$ be such that, for realizations $\mathbf{x}_E \neq \bar{\mathbf{x}}_E$,

$$
\begin{aligned}
&\pi_i\left(j, \mathbf{y}_{-i}, \mathbf{x}_{NE}, \mathbf{x}_E, \boldsymbol{\nu}, \varepsilon_{ij}\right) - \pi_i\left(j, \mathbf{y}_{-i}, \mathbf{x}_{NE}, \bar{\mathbf{x}}_E, \boldsymbol{\nu}, \varepsilon_{ij}\right) \\
=&\pi_i\left(k, \mathbf{y}_{-i}, \mathbf{x}_{NE}, \mathbf{x}_E, \boldsymbol{\nu}, \varepsilon_{ik}\right) - \pi_i\left(k, \mathbf{y}_{-i}, \mathbf{x}_{NE}, \bar{\mathbf{x}}_E, \boldsymbol{\nu}, \varepsilon_{ik}\right)
\end{aligned}
\tag{15}
$$

for $j \neq k$ and $\forall i \in \mathcal{N}$. In other words, the variation in payoffs induced by $\mathcal{X}_E$ does not vary with players' choices. Rearranging the previous equation, one obtains

$$
\begin{aligned}
&\pi_i\left(j, \mathbf{y}_{-i}, \mathbf{x}_{NE}, \mathbf{x}_E, \boldsymbol{\nu}, \varepsilon_{ij}\right) - \pi_i\left(k, \mathbf{y}_{-i}, \mathbf{x}_{NE}, \mathbf{x}_E, \boldsymbol{\nu}, \varepsilon_{ik}\right) \\
=&\pi_i\left(j, \mathbf{y}_{-i}, \mathbf{x}_{NE}, \bar{\mathbf{x}}_E, \boldsymbol{\nu}, \varepsilon_{ij}\right) - \pi_i\left(k, \mathbf{y}_{-i}, \mathbf{x}_{NE}, \bar{\mathbf{x}}_E, \boldsymbol{\nu}, \varepsilon_{ik}\right).
\end{aligned}
\tag{16}
$$

Since this equality holds $\forall i \in \mathcal{N}$, integrating over competitors' decisions implies that the difference

$$\pi_i^{\mathbf{P}}\left(j, \mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}}, \boldsymbol{\nu}, \varepsilon_{ij}\right) - \pi_i^{\mathbf{P}}\left(k, \mathbf{x}_{\mathrm{NE}}, \mathbf{x}_{\mathrm{E}}, \boldsymbol{\nu}, \varepsilon_{ik}\right) \tag{17}$$

does not depend on $\mathbf{x}_{\mathrm{E}}$. As a result, equilibrium choice probabilities do not depend on $\mathbf{x}_{\mathrm{E}}$ given $\mathbf{x}_{\mathrm{NE}}$ and $\boldsymbol{\nu}$.

Variables $\mathcal{X}_{\mathrm{E}}$ satisfying (15) could be some observable outcomes related to players' previous decisions that are realized before the game is played and are taken as given in the game that is studied by the researcher. Such variables are predetermined outcomes. The key for (15) to be satisfied is that, even if those previous decisions do affect players' current payoffs, this effect must be the same regardless of the decisions made by the players. For instance, in games of market entry, predetermined outcomes could be previously acquired fixed assets that are in use no matter whether the firm decides to enter or not, and are not modifiable in the game of interest.

Predetermined outcomes should therefore be informative about common knowledge state variables available to the players, including information that is unobservable to the econometrician and therefore satisfy the relevance condition. Since, from equation (17), equilibrium choice probabilities do not depend on such predetermined outcomes, the redundancy condition is also satisfied if these variables do not affect equilibrium selection (given $\mathcal{X}_{\mathrm{NE}}$ and $\mathcal{V}$).

Of course, the equilibrium selection being conditionally independent of the predetermined outcomes does put some restrictions on the type of variables that may be used as proxies. In particular, for predetermined outcomes to be valid, they should not be part of a multistage game where previous actions could be informative about players' decisions and equilibrium selection beyond $\mathcal{X}_{\mathrm{NE}}$ and $\mathcal{V}$. For example, predetermined outcomes allowing players to signal their willingness to take specific future actions would not be applicable in the current setting. As suggested by Van Damme (1989) and Ben-Porath (1992), these signals could alter the equilibrium selection and would therefore not satisfy the desired exclusion restriction. To the extent that one can make the economic argument that the outcomes used as proxies are not predictive of which equilibrium of the game is realized (once again, given $\mathcal{X}_{\mathrm{NE}}$ and $\mathcal{V}$), the conditions for valid proxies are satisfied. If one is able to convincingly argue that the proposed predetermined outcomes are not part of a multistage forward-looking decision process, they are unlikely to affect equilibrium selection (given $\mathcal{X}_{\mathrm{NE}}$ and $\mathcal{V}$). In such cases, provided that (15) holds, $\mathcal{X}_{\mathrm{E}}$ can be used as proxies.

Below is an example of predetermined outcomes that can be used as proxies for common knowledge unobserved heterogeneity in a game of market entry between two players. This example is a simplification of the model proposed by Gowrisankaran and Krainer (2011).

Example 4 (Potential locations in an entry game). Gowrisankaran and Krainer (2011) study a game of automatic teller machine (ATM) location and pricing

decisions. There are two types of firms, banks, and nonbanks, deciding whether or not to install an ATM on their premises and what level of fees to charge to their ATM users. Consumers choose where to withdraw money based on ATM locations and fees. Potential locations for ATMs are existing banks and retail establishments (mostly grocery and convenience stores).

Gowrisankaran and Krainer's (2011) estimation method assumes that the same equilibrium is played in different markets with identical observables. In their setting, a market is a rural county. Since rural areas are typically less densely populated than cities, different subsets of firms installing an ATM on their premises may be sustainable in equilibrium. The single equilibrium in the data assumption maintains that if banks are more likely than nonbanks to install an ATM in a given market, it must also be the case in all other markets that are similar according to common knowledge state variables.

The proposed method can be used to test the single equilibrium in the data assumption in this setting. One important form of unobserved heterogeneity that is worth taking into account when testing this assumption is consumers' preferences for banking services across markets. In fact, banks being more likely to install ATMs in a given market could either be due: (a) to consumers frequently visiting banking establishments; or (b) to the equilibrium in which banks are more likely to install ATMs being realized in this market.

Consider a simple version of the ATM location decision game (similar to an entry game). Let $i = 1$ be banks and $i = 2$ be nonbanks. A larger number of categories of players could be used as long as one is willing to assume that the conditional choice probability function is the same across players within a given category. Let $y_i = 1$ if player i installs an ATM; $y_i = 0$ otherwise. Suppose that, given $\mathbf{x}_{NE}$, there are two types of markets such that $|\mathscr{V}(\mathbf{x}_{NE})| = 2$. In markets with $\mathbf{v}^1$, consumers are frequent users of banking services and, therefore, they often visit banking establishments. In markets with $\mathbf{v}^0$, consumers use banking services less frequently. Since banks are more likely to install an ATM in markets with realization $\mathbf{v}^1$, $p_1(\mathbf{x}_{NE}, \mathbf{v}^1) > p_1(\mathbf{x}_{NE}, \mathbf{v}^0)$. Notice that one could allow for up to $2^2 = 4$ elements in $\mathscr{V}(\mathbf{x}_{NE})$ (provided that there is enough variation in the support $\mathscr{X}_E$), but considering $|\mathscr{V}(\mathbf{x}_{NE})| = 2$ in this setting shows that even binary common knowledge unobserved heterogeneity may capture relevant variation in the data.

Since Gowrisankaran and Krainer (2011) consider a game of ATM installation taking potential ATM locations as given, one could use information about banks' and nonbanks' establishments within a market (e.g., $\mathscr{X}_E$ could include two variables corresponding to the number of each players' potential locations) as predetermined outcomes that are valid proxies for this common knowledge unobserved heterogeneity. More precisely, banks' and nonbanks' establishments are the only potential locations for the game of interest. These predetermined outcomes have been decided prior to the game of interest and their effects on banks' and nonbanks' payoffs are the same regardless whether they decide to install ATMs at these locations. Moreover, since these potential ATM locations are taken as given, they cannot be modified in the ATM locations game. In that

sense, they satisfy (15). In order to justify the redundancy of these predetermined outcomes, it remains to argue that potential locations do not affect equilibrium selection beyond $\mathcal{X}_{NE}$ and $\mathcal{V}$. In fact, it is unlikely that banks and nonbanks chose the location of their premises – in many cases, before the technology of ATMs was available – in order to signal they will install an ATM in a way that would affect equilibrium selection. Nonetheless, this proxy should reflect consumers' preferences for banking services: observing relatively more bank establishments in a market is potentially associated with consumers visiting banking establishments more frequently, which in itself should affect ATM location decisions. This correlation between the predetermined outcomes and preferences for banking services allows one to control for such common knowledge unobserved heterogeneity when testing the single equilibrium in the data assumption.

Predetermined outcomes as the ones described above also have the advantage of not requiring the researcher to know the source of unobserved heterogeneity one would like to control for. There is however one caveat: by definition, since predetermined outcomes are realized before the game of interest, they are more suitable to capture time-invariant unobserved heterogeneity. In many examples, including Example 4, time-invariant unobserved heterogeneity is indeed a relevant source of unobserved heterogeneity one would like to control for when testing the degenerate equilibrium selection assumption.

4.5. Identifying the Number of Components

The identification result in this paper holds for a given $|\mathcal{V}(\mathbf{x}_{NE})|$. It is worth noting that $|\mathcal{V}(\mathbf{x}_{NE})|$ is actually identifiable. Let $\Delta\mathbf{P}(\mathbf{x}_{NE})$ be the $(|\mathcal{X}_E|-1) \times (J+1)^N$ matrix with element (i,j) given by $p\left(\mathbf{y}^j\big|\mathbf{x}_{NE},\mathbf{x}_E^i\right) - p\left(\mathbf{y}^j\big|\mathbf{x}_{NE},\mathbf{x}_E^0\right)$, for $i=1,\ldots,|\mathcal{X}_E|-1$ and $j=1,\ldots,(J+1)^N$, i.e.,

$$\begin{bmatrix} p\left(\mathbf{y}^1\big|\mathbf{x}_{NE},\mathbf{x}_E^1\right)-p\left(\mathbf{y}^1\big|\mathbf{x}_{NE},\mathbf{x}_E^0\right) & \cdots & p\left(\mathbf{y}^{(J+1)^N}\big|\mathbf{x}_{NE},\mathbf{x}_E^1\right)-p\left(\mathbf{y}^{(J+1)^N}\big|\mathbf{x}_{NE},\mathbf{x}_E^0\right) \\ \cdots & \cdots & \cdots \\ p\left(\mathbf{y}^1\big|\mathbf{x}_{NE},\mathbf{x}_E^{|\mathcal{X}_E|-1}\right)-p\left(\mathbf{y}^1\big|\mathbf{x}_{NE},\mathbf{x}_E^0\right) & \cdots & p\left(\mathbf{y}^{(J+1)^N}\big|\mathbf{x}_{NE},\mathbf{x}_E^{|\mathcal{X}_E|-1}\right)-p\left(\mathbf{y}^{(J+1)^N}\big|\mathbf{x}_{NE},\mathbf{x}_E^0\right) \end{bmatrix}.$$

(18)

As it is stated in Lemma 1, for each $\mathbf{x}_{NE} \in \mathcal{X}_{NE}$, $|\mathcal{V}(\mathbf{x}_{NE})|$ is identified through the rank of the matrix $\Delta\mathbf{P}(\mathbf{x}_{NE})$.

LEMMA 1 (Identification of $|\mathcal{V}(\mathbf{x}_{NE})|$). *Under Assumptions 1 to 5, for each* $\mathbf{x}_{NE} \in \mathcal{X}_{NE}$, $|\mathcal{V}(\mathbf{x}_{NE})| = \text{rank}\{\Delta\mathbf{P}(\mathbf{x}_{NE})\}+1$.

Proof. This result follows from Henry et al. (2014), Lem. 2, p. 138). Details are in Appendix A.1. □

Since the elements in matrix $\Delta\mathbf{P}(\mathbf{x}_{NE})$ are constructed from point identified probabilities, its column rank is at most $(J+1)^N-1$. Therefore, since the rank of a matrix is bounded by the minimum number of its rows and columns, the

finite mixture representation restricts $|\mathscr{V}(\mathbf{x}_{\mathrm{NE}})|$ to be at most $\min\left\{(J+1)^N, |\mathscr{X}_{\mathrm{E}}|\right\}$, which corresponds to the cardinality of the support condition stated in Assumption 5. If players have symmetric choice probabilities, some of the $(J+1)^N$ columns are necessarily identical, such that the column rank is at most $J+1-1=J$. This is the reason why the bound stated in Example 3 is $\min\{J+1, |\mathscr{X}_{\mathrm{E}}|\}$.

An appealing by-product of Lemma 1 is that it can be used to test the hypothesis of the absence of payoff-relevant unobserved heterogeneity while remaining agnostic about the number of equilibria realized in the data. To the best of my knowledge, such a test is not available in the literature. In fact, from Lemma 1, $|\mathscr{V}(\mathbf{x}_{\mathrm{NE}})|=1$ if and only if $\mathrm{rank}\{\mathbf{\Delta P}(\mathbf{x}_{\mathrm{NE}})\}=0$, i.e., all the elements of the matrix in (18) are equal to 0. In other words, testing that all the elements of this matrix are jointly equal to 0 amounts to testing the absence of payoff-relevant unobserved heterogeneity. The test proposed in the current paper is a test of the degenerate equilibrium selection assumption that allows for correlation between players' decisions due to common knowledge unobserved heterogeneity. In that sense, it requires $|\mathscr{V}(\mathbf{x}_{\mathrm{NE}})| \geq 2$. Nonetheless, if one cannot reject that all the elements in (18) are jointly equal to 0, then one may interpret this finding as evidence against the presence of (discrete) payoff-relevant unobserved heterogeneity and can apply other tests available in the literature that assume away such unobserved heterogeneity (e.g., de Paula and Tang, 2012).

4.6. Constructing the Identified Set

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

The main intuition behind the identification argument combines results from Henry et al. (2014) and Kasahara and Shimotsu (2014). Let $\mathscr{N}^\xi$ denote some subset of $\mathscr{N}$ such that $\left|\mathscr{N}^\xi\right| = N^\xi \leq N$. Given the identifying restrictions stated above, one can write the unknown probabilities $p\left(\mathbf{y}^\xi \mid \mathbf{v}\right)$ $\forall \mathbf{y}^\xi \in \mathscr{Y}^{N^\xi}$, $\forall \mathscr{N}^\xi \subseteq \mathscr{N}$, $\forall \mathbf{v} \in \mathscr{V}$ and the mixing weights $\gamma\left(\mathbf{v} \mid \mathbf{x}_{\mathrm{E}}\right)$ $\forall \mathbf{x}_{\mathrm{E}} \in \mathscr{X}_{\mathrm{E}}$, $\forall \mathbf{v} \in \mathscr{V}$ as functions of a vector of parameters $\boldsymbol{\theta}$ to be defined below. Then, one can construct the identified set Θ_{I} such that $\boldsymbol{\theta}$ belongs to Θ_{I} if and only if the data can be rationalized by the finite mixture over $\mathscr{V}$ corresponding to $\boldsymbol{\theta}$ and the degenerate equilibrium selection, i.e., $|\mathscr{T}^*(\mathbf{v})| = 1$ $\forall \mathbf{v} \in \mathscr{V}$. For a given $\mathbf{x}_{\mathrm{NE}}$, $\Theta_{\mathrm{I}} \neq \emptyset$ is therefore a testable implication of the single equilibrium in the data assumption. There are two types of restrictions defining the identified set.

1. The distribution of players' decisions can be written as a finite mixture over $\mathscr{V}$, i.e., the unknown $\mathbf{v}$-specific probabilities and the mixing weights are proper probability mass functions.
2. The distributions of players' decisions satisfy the conditions implied by a single equilibrium being realized in the data, i.e., players' decisions are independent after conditioning on all common knowledge information.

While the first type of restrictions are similar to the ones provided in Henry et al. (2014), the second type of restrictions are specific to testing the degenerate equilibrium selection assumption and can be stated using an identification result from Kasahara and Shimotsu (2014).

Some notation must be introduced before stating formal results. First, the matrix $\mathbf{\Delta P}$ is defined as in (18). Let $\mathbf{\Delta \tilde{P}}$ be a $(|\mathcal{V}|-1) \times (|\mathcal{V}|-1)$ matrix obtained by deleting some rows and some columns of $\mathbf{\Delta P}$. For subsets $\tilde{\mathscr{X}_E}$ and $\widetilde{\mathscr{Y}^N}$ satisfying Assumptions 5(v) and 5(vi), the matrix $\mathbf{\Delta \tilde{P}}$ is invertible. Let $\mathbf{\Delta \tilde{p}}^r(\mathbf{x}_E)'$ be the row of $\mathbf{\Delta P}$ corresponding to $\mathbf{x}_E$ after dropping the columns that are not in $\mathbf{\Delta \tilde{P}}$. Similarly, let $\mathbf{\Delta \tilde{p}}^c(\mathbf{y})$ be the column of $\mathbf{\Delta P}$ corresponding to $\mathbf{y}$ after dropping the rows that are not in $\mathbf{\Delta \tilde{P}}$. Similar matrices and vectors $\mathbf{\Delta \tilde{P}}^\zeta$, $\mathbf{\Delta \tilde{p}}^{\zeta,r}(\mathbf{x}_E)$ and $\mathbf{\Delta \tilde{p}}^{\zeta,c}(\mathbf{y}^\zeta)$ can be defined for any subset $\mathcal{N}^\zeta$. Furthermore, denote

$$\boldsymbol{\gamma}(\mathbf{x}_E) \equiv \left[\gamma\left(\boldsymbol{v}^1 \big| \mathbf{x}_E\right), \ldots, \gamma\left(\boldsymbol{v}^{|\mathcal{V}|-1} \big| \mathbf{x}_E\right) \right]'. \tag{19}$$

While Henry et al. (2014) study the case where $\mathbf{y}$ is a scalar, their results directly extend to $p\left(\mathbf{y}^\zeta \big| \boldsymbol{v}\right) \; \forall \mathbf{y}^\zeta \in \mathscr{Y}^{N^\zeta}$, $\boldsymbol{v} \in \mathcal{V}$ for any subset $\mathcal{N}^\zeta$. Proposition 1 states that these probabilities and $\gamma(\boldsymbol{v}|\mathbf{x}_E) \; \forall \mathbf{x}_E \in \mathscr{X}_E$, $\boldsymbol{v} \in \mathcal{V}$ can be written as functions of $|\mathcal{V}| \times (|\mathcal{V}|-1)$ parameters stored in the $(|\mathcal{V}|-1) \times 1$ vector $\boldsymbol{\phi}$ and the $(|\mathcal{V}|-1) \times (|\mathcal{V}|-1)$ matrix $\boldsymbol{\Upsilon}$ defined as[18]

$$\boldsymbol{\phi} \equiv \left[\gamma\left(\boldsymbol{v}^1 \big| \mathbf{x}_E^0\right), \ldots, \gamma\left(\boldsymbol{v}^{|\mathcal{V}|-1} \big| \mathbf{x}_E^0\right) \right]' \tag{20}$$

and

$$\boldsymbol{\Upsilon} \equiv \begin{bmatrix} \Delta\gamma\left(\boldsymbol{v}^1 \big| \tilde{\mathbf{x}}_E^1\right) & \ldots & \Delta\gamma\left(\boldsymbol{v}^1 \big| \tilde{\mathbf{x}}_E^{|\mathcal{V}|-1}\right) \\ \ldots & \ldots & \ldots \\ \Delta\gamma\left(\boldsymbol{v}^{|\mathcal{V}|-1} \big| \tilde{\mathbf{x}}_E^1\right) & \ldots & \Delta\gamma\left(\boldsymbol{v}^{|\mathcal{V}|-1} \big| \tilde{\mathbf{x}}_E^{|\mathcal{V}|-1}\right) \end{bmatrix}, \tag{21}$$

where, $\Delta\gamma\left(\boldsymbol{v}|\tilde{\mathbf{x}}_E\right) \equiv \gamma\left(\boldsymbol{v}|\tilde{\mathbf{x}}_E\right) - \gamma\left(\boldsymbol{v}|\mathbf{x}_E^0\right)$. From Assumption 5(v), $\boldsymbol{\Upsilon}$ is invertible. The vector $\boldsymbol{\theta}$ collects all the elements of $\boldsymbol{\phi}$ and $\boldsymbol{\Upsilon}$.

PROPOSITION 1 (Finite mixture probabilities). *For an arbitrary $|\mathcal{V}|$, if Assumptions 1–5 are satisfied, then, $\forall \mathbf{y}^\zeta \in \mathscr{Y}^{N^\zeta}$, $\forall \mathcal{N}^\zeta \subseteq \mathcal{N}$ and $l = 0, 1, \ldots, |\mathcal{V}|-1$:*

$$p\left(\mathbf{y}^\zeta \big| \boldsymbol{v}^l\right) = p\left(\mathbf{y}^\zeta \big| \mathbf{x}_E^0\right) + [\mathbf{e}_l - \boldsymbol{\phi}]' \, \boldsymbol{\Upsilon}'^{-1} \mathbf{\Delta \tilde{p}}^{\zeta,c}\left(\mathbf{y}^\zeta\right), \tag{22}$$

where $\mathbf{e}_l$ is the $(|\mathcal{V}|-1) \times 1$ unit vector with lth row equal to 1 and $\mathbf{e}_0 \equiv \mathbf{0}$. Moreover, $\forall \mathbf{x}_E \in \mathscr{X}_E$:

$$\boldsymbol{\gamma}(\mathbf{x}_E) = \boldsymbol{\phi} + \boldsymbol{\Upsilon} \left[\mathbf{\Delta \tilde{P}}'\right]^{-1} \mathbf{\Delta \tilde{p}}^r(\mathbf{x}_E). \tag{23}$$

---

[18]Once again, more rigorous notation would write $\boldsymbol{\phi}(\mathbf{x}_{NE})$ and $\boldsymbol{\Upsilon}(\mathbf{x}_{NE})$, but the argument is dropped to alleviate notation.

Proof. The result follows from arguments similar to Henry et al. (2014), Sects. 3 and 4.2). Details are available in Appendix A.2.   □

The expressions of the finite mixture probabilities in Proposition 1 are used to define the identified set Θ_I of all the values of $\boldsymbol{\phi}, \boldsymbol{\Upsilon}$ that rationalize the distribution of players' decisions via a finite mixture over $\mathcal{V}$ and a single equilibrium being realized in the data. The result is formally stated in Proposition 2. The following notation is used. Let $\mathcal{N}_n^\zeta$ denote the subset of $\mathcal{N}$ containing exactly players $i = 1, \ldots, n$. Furthermore, let $\mathbf{P}_n^\zeta(\boldsymbol{v}^l)$ be the $(J+1)^n \times (J+1)$ matrix of joint probabilities conditional on $\boldsymbol{v}^l$ over the n-dimensional vector containing the decisions of players $1, \ldots, n$ – i.e., $\mathbf{y}_n^\zeta \in \{\mathbf{y}_n^{\zeta,1}, \ldots, \mathbf{y}_n^{\zeta,(J+1)^n}\}$) – and player $n+1$'s decision – i.e., $y_{n+1} \in \{y_{n+1}^1, \ldots, y_{n+1}^{J+1}\}$ – such that

$$\mathbf{P}_n^\zeta(\boldsymbol{v}^l) = \begin{bmatrix} p\left(\mathbf{y}_n^{\zeta,1}, y_{n+1}^1 \middle| \boldsymbol{v}^l\right) & \cdots & p\left(\mathbf{y}_n^{\zeta,1}, y_{n+1}^{J+1} \middle| \boldsymbol{v}^l\right) \\ \cdots & \cdots & \cdots \\ p\left(\mathbf{y}_n^{\zeta,(J+1)^n}, y_{n+1}^1 \middle| \boldsymbol{v}^l\right) & \cdots & p\left(\mathbf{y}_n^{\zeta,(J+1)^n}, y_{n+1}^{J+1} \middle| \boldsymbol{v}^l\right) \end{bmatrix}. \quad (24)$$

Using Proposition 1, one can write the (i,j)th element of $\mathbf{P}_n^\zeta(\boldsymbol{v}^l)$ for $l \in \{0, \ldots, |\mathcal{V}|-1\}$ as

$$p\left(\mathbf{y}_n^{\zeta,i}, y_{n+1}^j \middle| \boldsymbol{v}^l\right) = p\left(\mathbf{y}_n^{\zeta,i}, y_{n+1}^j \middle| \mathbf{x}_E^0\right) + [\mathbf{e}_l - \boldsymbol{\phi}]' \boldsymbol{\Upsilon}'^{-1} \Delta \tilde{\mathbf{p}}^{\zeta,c}\left(\mathbf{y}_n^{\zeta,i}, y_{n+1}^j\right). \quad (25)$$

PROPOSITION 2 (Identified set). *For an arbitrary $|\mathcal{V}|$, under Assumptions 1–5, $\boldsymbol{\theta} \in \Theta_I$ if and only if:*

(i) $0 \leq \boldsymbol{\phi} + \boldsymbol{\Upsilon}\mathbf{a}$ *and* $\mathbf{1}'[\boldsymbol{\phi} + \boldsymbol{\Upsilon}\mathbf{a}] < 1$, $\forall \mathbf{a} \in \mathcal{A}$, *where $\mathcal{A}$ is the set of extreme points of the convex hull of the range of the function* $\mathbf{x}_E \mapsto \left[\Delta \tilde{\mathbf{P}}'\right]^{-1} \Delta \tilde{\mathbf{p}}^r(\mathbf{x}_E)$.

(ii) $[\mathbf{e}_l - \boldsymbol{\phi}]' \boldsymbol{\Upsilon}'^{-1}\mathbf{b} \leq 1$, $\forall l \in \{0, \ldots, |\mathcal{V}|-1\}$, $\forall \mathbf{b} \in \mathcal{B}$, *where $\mathcal{B}$ is the union of the ranges of* $\mathbf{y} \mapsto \dfrac{-\Delta \tilde{\mathbf{p}}^c(\mathbf{y})}{P\left(\mathbf{y}|\mathbf{x}_E^0\right)}$ *and* $\mathbf{y} \mapsto \dfrac{\Delta \tilde{\mathbf{p}}^c(\mathbf{y})}{1-P\left(\mathbf{y}|\mathbf{x}_E^0\right)}$.

(iii) Rank$\left\{\mathbf{P}_n^\zeta(\boldsymbol{v}^l)\right\} = 1$, $\forall n \in \{1, \ldots, N-1\}$ *and* $\forall l \in \{0, \ldots, |\mathcal{V}|-1\}$, *where the (i,j)th element of $\mathbf{P}_n^\zeta(\boldsymbol{v}^l)$ is written as in (25).*

Proof. See Appendix A.3.   □

The conditions defining Θ_I in Proposition 2 are interpreted as follows. Parts (i) and (ii) guarantee that the probabilities defining the finite mixture over the common knowledge unobserved heterogeneity properly belong to the unit interval. These inequalities are the same as in Henry et al. (2014), Sects. 3 and 4.2). More precisely, part (i) is equivalent to $\gamma\left(\boldsymbol{v}^l \middle| \mathbf{x}_E\right) \geq 0$ for $l = 1, \ldots, |\mathcal{V}|-1$ and $\sum_{l=1}^{|\mathcal{V}|-1} \gamma\left(\boldsymbol{v}^l \middle| \mathbf{x}_E\right) < 1$ $\forall \mathbf{x}_E \in \mathcal{X}_E$. Furthermore, part (ii) states that $0 \leq p\left(\mathbf{y}|\boldsymbol{v}\right) \leq 1$ $\forall \mathbf{y} \in \mathcal{Y}^N$ and $\forall \boldsymbol{v} \in \mathcal{V}$. Part (iii) follows from the degenerate equilibrium selection assumption. The fact that this condition can be characterized via the rank of matrices $\mathbf{P}_n^\zeta(\boldsymbol{v}^l)$ being equal to 1 follows from Kasahara and Shimotsu (2014,

Prop. 4). It is a consequence of the conditional independence of players' decisions when a single equilibrium is realized in the data.

Several remarks are worth making about the identified set.

Remark 1 (Sharp identified set). The identified set Θ_I is sharp in the following sense. Given $\mathcal{X}_{NE}$, if the true data generating process corresponds to a single equilibrium being realized $\forall v \in \mathcal{V}$, then the value of θ corresponding to this true data generating process belongs to Θ_I and Θ_I cannot be empty. Conversely, each $\theta \in \Theta_I$ corresponds to a finite mixture over $\mathcal{V}$ which rationalizes the joint distribution of players' decisions conditional on $\mathcal{X}_{NE}$ through a single equilibrium at each realization of $\mathcal{V}$. As a result, if Θ_I is not empty, the data can be rationalized by the degenerate equilibrium selection and the finite mixture unobserved heterogeneity assumptions.

Remark 2 (Equality restrictions). The conditions in part (iii) from Proposition 2 are equality restrictions and, therefore, drastically shrink the identified set. In fact, as it is stated in Corollary 1, they shrink the identified set to a singleton if $|\mathcal{V}| = 2$ and the identified set is nonempty.

Remark 3 (Separately verifying finite mixture over unobserved heterogeneity). By the nature of its defining restrictions, one can check whether Θ_I being empty is due to a violation of the single equilibrium in the data assumption or the finite mixture representation of the unobserved heterogeneity. Here, the identifying restrictions in Assumption 5 are key. They allow one to identify $p(\mathbf{y}|v)$'s, regardless of the number of equilibria realized at each $v \in \mathcal{V}$, using conditions (i) and (ii) in Proposition 2. In the Monte Carlo simulations presented in Section 6, the finite mixture restrictions are separately tested in order to assess what fraction of test rejections are actually due the finite mixture representation being rejected when it is true.

Remark 4 (Number of players and actions). A nice consequence of Proposition 2 is that increasing the number of players and/or the number of actions in the choice set does not affect the dimension of Θ, which is entirely driven by $|\mathcal{V}|$. In that sense, increasing N and/or J only adds more conditions that further restrict Θ_I.

Remark 5 (Separating the finite mixtures). Variation in $\mathcal{X}_E$ restricts the set of θ satisfying the conditions implied by the mixture over $\mathcal{V}$. However, variation in $\mathcal{X}_E$ does not affect the conditions related to the single equilibrium in the data assumption. In fact, $\mathcal{X}_E$ is excluded from the equilibrium choice probabilities and the equilibrium selection mechanism. In that sense, the proxy variables are key to separate the two layers of finite mixtures: the one over the multiple equilibria and the one over the unobserved heterogeneity.

4.7. The Special Case of a Mixture over Two Components

It is easy to represent graphically the identified set for $|\mathcal{V}| = 2$ (for any N and J). In that case, $\gamma(\mathbf{x}_E) = \gamma(v^1|\mathbf{x}_E)$ and

$$p(\mathbf{y}|\mathbf{x}_E) = p(\mathbf{y}|\boldsymbol{v}^0) + \left[p(\mathbf{y}|\boldsymbol{v}^1) - p(\mathbf{y}|\boldsymbol{v}^0)\right]\gamma(\mathbf{x}_E). \tag{26}$$

There are two parameters in $\boldsymbol{\theta}$, i.e., ϕ and Υ, which are respectively given by the first elements of (20) and (21). For $|\mathscr{V}| = 2$, Proposition 1 implies that $\forall \mathbf{y} \in \mathscr{Y}^N$:

$$p(\mathbf{y}|\boldsymbol{v}^0) = p(\mathbf{y}|\mathbf{x}_E^0) - \frac{\phi}{\Upsilon}\left[p(\mathbf{y}|\tilde{\mathbf{x}}_E^1) - p(\mathbf{y}|\mathbf{x}_E^0)\right], \tag{27}$$

$$p(\mathbf{y}|\boldsymbol{v}^1) = p(\mathbf{y}|\mathbf{x}_E^0) + \frac{1-\phi}{\Upsilon}\left[p\left(\mathbf{y}|\tilde{\mathbf{x}}_E^1\right) - p(\mathbf{y}|\mathbf{x}_E^0)\right]. \tag{28}$$

Moreover, $\forall\, \mathbf{x}_E \in \mathscr{X}_E$:

$$\gamma(\mathbf{x}_E) = \phi + \Upsilon \frac{p\left(\tilde{\mathbf{y}}^1 \middle| \mathbf{x}_E\right) - p\left(\tilde{\mathbf{y}}^1 \middle| \mathbf{x}_E^0\right)}{p\left(\tilde{\mathbf{y}}^1 \middle| \tilde{\mathbf{x}}_E^1\right) - p\left(\tilde{\mathbf{y}}^1 \middle| \mathbf{x}_E^0\right)}. \tag{29}$$

For $|\mathscr{V}| = 2$, it is useful to write the conditions defining Θ_I in terms of the reparameterization $-\phi/\Upsilon$, $(1-\phi)/\Upsilon$ and the following functions:

$$L_0(\mathbf{y}) \equiv \frac{-p(\mathbf{y}|\mathbf{x}_E^0)}{p\left(\mathbf{y}|\tilde{\mathbf{x}}_E^1\right) - p(\mathbf{y}|\mathbf{x}_E^0)}; \, L_1(\mathbf{y}) \equiv \frac{1 - p(\mathbf{y}|\mathbf{x}_E^0)}{p\left(\mathbf{y}|\tilde{\mathbf{x}}_E^1\right) - p(\mathbf{y}|\mathbf{x}_E^0)};$$

$$\text{and } Q(\mathbf{x}_E) \equiv \frac{p\left(\tilde{\mathbf{y}}^1 \middle| \mathbf{x}_E\right) - p\left(\tilde{\mathbf{y}}^1 \middle| \mathbf{x}_E^0\right)}{p\left(\tilde{\mathbf{y}}^1 \middle| \tilde{\mathbf{x}}_E^1\right) - p\left(\tilde{\mathbf{y}}^1 \middle| \mathbf{x}_E^0\right)}. \tag{30}$$

Notice that $Q(\mathbf{x}_E)$ can be rewritten as $\left[\gamma(\mathbf{x}_E) - \gamma(\mathbf{x}_E^0)\right] / \left[\gamma(\tilde{\mathbf{x}}_E^1) - \gamma(\mathbf{x}_E^0)\right]$ and therefore does not depend on $\tilde{\mathbf{y}}^1$.

Consider the case where $p\left(\mathbf{y}|\tilde{\mathbf{x}}_E^1\right) - p(\mathbf{y}|\mathbf{x}_E^0) > 0$. Part (i) in Proposition 2 simplifies to $0 < \gamma(\mathbf{x}_E) < 1 \,\, \forall \mathbf{x}_E \in \mathscr{X}_E$. For a given $\mathbf{x}_E$, one can write,

$$\gamma(\mathbf{x}_E) > 0 \Rightarrow \frac{-\phi}{\Upsilon} < Q(\mathbf{x}_E) \text{ and} \tag{31}$$

$$\gamma(\mathbf{x}_E) < 1 \Rightarrow \frac{1-\phi}{\Upsilon} > Q(\mathbf{x}_E). \tag{32}$$

Moreover, part (ii) in Proposition 2 simplifies to $0 \le p\left(\mathbf{y}|\boldsymbol{v}^l\right) \le 1 \,\, \forall \mathbf{y} \in \mathscr{Y}^N$ and $l = 0, 1$. For a given $\mathbf{y} \in \mathscr{Y}^N$,

$$0 \le p\left(\mathbf{y}|\boldsymbol{v}^0\right) \le 1 \Rightarrow L_0(\mathbf{y}) \le \frac{-\phi}{\Upsilon} \le L_1(\mathbf{y}) \text{ and} \tag{33}$$

$$0 \le p\left(\mathbf{y}|\boldsymbol{v}^1\right) \le 1 \Rightarrow L_0(\mathbf{y}) \le \frac{1-\phi}{\Upsilon} \le L_1(\mathbf{y}). \tag{34}$$

Considering $p\left(\mathbf{y}|\tilde{\mathbf{x}}_E^1\right) - p\left(\mathbf{y}|\mathbf{x}_E^0\right) > 0$ implies that $L_0(\mathbf{y}) \le Q(\mathbf{x}_E) \le L_1(\mathbf{y})$. Therefore the upper bound on $-\phi/\Upsilon$ in (33) is satisfied whenever (31) holds. Similarly, the lower bound on $(1-\phi)/\Upsilon$ in (34) is satisfied whenever (32) holds. Inequalities (31)–(34) can therefore be summarized as

$$L_0(\mathbf{y}) \le \frac{-\phi}{\Upsilon} < Q(\mathbf{x}_E) \text{ and } Q(\mathbf{x}_E) < \frac{1-\phi}{\Upsilon} \le L_1(\mathbf{y}). \tag{35}$$

A similar argument for the case where $p\left(\mathbf{y}|\tilde{\mathbf{x}}_E^1\right) - p\left(\mathbf{y}|\mathbf{x}_E^0\right) < 0$, which in turn implies that $L_1(\mathbf{y}) \le Q(\mathbf{x}_E) \le L_0(\mathbf{y})$, gives

$$L_1(\mathbf{y}) \le \frac{-\phi}{\Upsilon} < Q(\mathbf{x}_E) \text{ and } Q(\mathbf{x}_E) < \frac{1-\phi}{\Upsilon} \le L_0(\mathbf{y}). \tag{36}$$

Since these inequalities must hold for all $\mathbf{y} \in \mathscr{Y}^N$ and all $\mathbf{x}_E \in \mathscr{X}_E$, the restrictions on $\boldsymbol{\theta}$ coming from Proposition 2(i) and (ii) boil down to

$$\max_{\mathbf{y} \in \mathscr{Y}^N} \{\min\{L_0(\mathbf{y}), L_1(\mathbf{y})\}\} \le \frac{-\phi}{\Upsilon} < \min_{\mathbf{x}_E \in \mathscr{X}_E} \{Q(\mathbf{x}_E)\} \text{ and} \tag{37}$$

$$\max_{\mathbf{x}_E \in \mathscr{X}_E} \{Q(\mathbf{x}_E)\} < \frac{1-\phi}{\Upsilon} \le \min_{\mathbf{y} \in \mathscr{Y}^N} \{\max\{L_0(\mathbf{y}), L_1(\mathbf{y})\}\}. \tag{38}$$

When $|\mathscr{V}| = 2$, part (iii) from Proposition 2 is satisfied by at most one possible couple $-\phi/\Upsilon$, $(1-\phi)/\Upsilon$. To see this, notice that for $n = 1$, $\mathbf{P}_n^\zeta(\boldsymbol{v}^l)$ is simply the matrix of the joint distribution of the decisions from players 1 and 2. A necessary condition for rank $\{\mathbf{P}_n^\zeta(\boldsymbol{v}^l)\} = 1$ is that all minors of order 2 computed from $\mathbf{P}_n^\zeta(\boldsymbol{v}^l)$ must be equal to 0. As a result, using y_i^1, y_i^2 to denote two arbitrary actions from player $i = 1, 2$, the following equation must hold for $l = 0, 1$:

$$p\left(y_1^1, y_2^1 | \boldsymbol{v}^l\right) p\left(y_1^2, y_2^2 | \boldsymbol{v}^l\right) - p\left(y_1^1, y_2^2 | \boldsymbol{v}^l\right) p\left(y_1^2, y_2^1 | \boldsymbol{v}^l\right) = 0. \tag{39}$$

Using Proposition 1, (39) defines a quadratic equation in $-\phi/\Upsilon$ for $l = 0$ and another quadratic equation in $(1-\phi)/\Upsilon$ for $l = 1$. These quadratic equations each have at most two solutions. Denote these solutions $(-\phi/\Upsilon)^*$ and $(-\phi/\Upsilon)^{**}$ for $l = 0$; $((1-\phi)/\Upsilon)^*$ and $((1-\phi)/\Upsilon)^{**}$ for $l = 1$. There are therefore at most four possible candidates of couples $-\phi/\Upsilon$, $(1-\phi)/\Upsilon$ that can satisfy the conditions in Proposition 2(iii). Notice that, since the quadratic equations in (39) have the same coefficients, the solutions are the same for both equations, i.e., $(-\phi/\Upsilon)^* = ((1-\phi)/\Upsilon)^*$ and $(-\phi/\Upsilon)^{**} = ((1-\phi)/\Upsilon)^{**}$. However, since $-\phi/\Upsilon \ne (1-\phi)/\Upsilon$, 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 $(1-\phi)/\Upsilon > 0$. Let $(-\phi/\Upsilon)^*$, $((1-\phi)/\Upsilon)^{**}$ be this couple.

The discussion above is summarized in Figure 2. The shaded area corresponds to $\boldsymbol{\theta}$'s that satisfy conditions (i) and (ii) from Proposition 2. The couple $(-\phi/\Upsilon)^*$,
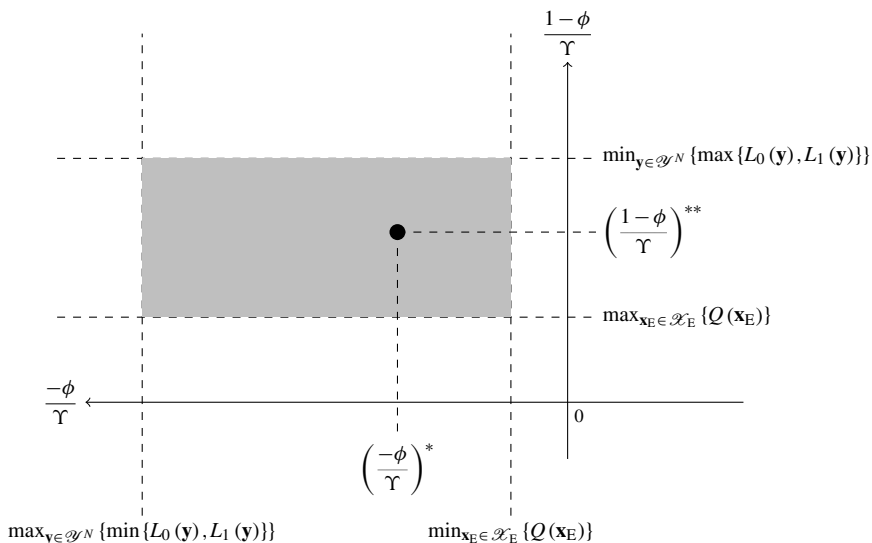
FIGURE 2. Graphical representation of the identified set for the case $|\mathscr{V}| = 2$.

$((1 - \phi)/\Upsilon)^{**}$, if it exists, is the only one that satisfies conditions (iii). As a result, $\Theta_I \neq \emptyset$ if and only if this point falls in the shaded area.

As it is made obvious in Figure 2, part (iii) of Proposition 2 shrinks the identified set to a singleton. This result is formally stated in Corollary 1.

COROLLARY 1 (Singleton nonempty Θ_I). *If $|\mathscr{V}| = 2$ and $\Theta_I \neq \emptyset$, then Θ_I is a singleton.*

Proof. The proof directly follows from noting that there is at most one couple $-\phi/\Upsilon, (1 - \phi)/\Upsilon$ that satisfies conditions (iii) in Proposition 2. If it exists, this couple defines a system of two equations in two unknowns which leads to at most one $\boldsymbol{\theta} \in \Theta_I$. $\square$

Corollary 1 implies that if the data can be rationalized by a single equilibrium and a finite mixture over $|\mathscr{V}| = 2$ values of $\boldsymbol{v}$, the corresponding $\boldsymbol{\theta}$ is point identified (for any N and J). One way to interpret Corollary 1 is that, if one is willing to assume that the degenerate equilibrium selection assumption holds, players' decisions are independent given $\mathcal{V}$ and the finite mixture over $\mathcal{V}$ is point-identified from variation in the proxy for $\mathcal{V}$ when $|\mathscr{V}| = 2$. This result holds for any N and J.

Remark 6 (Point identification and relation to the literature). The result in Corollary 1 leverages both the variation in the proxy for $\mathcal{V}$ and the conditional independence of $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$ given $\mathcal{V}$. In that sense, the fact that it delivers point identification instead of set identification as in Henry et al. (2014) is not too

surprising. Furthermore, applying the results from Hall and Zhou (2003) to the current setting would imply that point identification of the mixture over $\mathcal{V}$ should require $N \geq 3$. However, Hall and Zhou (2003) only leverage the conditional independence of $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$ given $\mathcal{V}$. The mixture being point-identified even if $N = 2$ in Corollary 1 is due to the identifying power of the proxy for $\mathcal{V}$.

Remark 7 (Point identification with $|\mathcal{V}| > 2$). The point identification result in Corollary 1 does not directly extend to $|\mathcal{V}| > 2$. The key feature of the $|\mathcal{V}| = 2$ case is that the conditions in part (iii) of Proposition 2 define equality constraints on the reparameterization $-\phi/\Upsilon$, $(1 - \phi)/\Upsilon$ which can be used to construct a system of two equations that are linear in the two unknowns ϕ and Υ. In the case $|\mathcal{V}| > 2$, this reparameterization becomes $[\mathbf{e}_l - \boldsymbol{\phi}]' \, \boldsymbol{\Upsilon}'^{-1}$ for $l \in \{0, 1, \ldots, |\mathcal{V}| - 1\}$. Even in cases where the conditions in part (iii) from Proposition 2 would be satisfied by a unique value of the reparameterized parameters, the resulting system of $|\mathcal{V}| \times (|\mathcal{V}| - 1)$ equations is not linear in the $|\mathcal{V}| \times (|\mathcal{V}| - 1)$ elements of $\boldsymbol{\theta}$. There is therefore no guarantee that a unique value of $\boldsymbol{\theta}$ satisfies the conditions defining Θ_{I}.

5. STATISTICAL TEST

5.1. Testable Implications

The null hypothesis $\mathrm{H}_0 : |\mathscr{T}^*(\mathbf{x}, \boldsymbol{v})| = 1, \forall \boldsymbol{v} \in \mathscr{V}(\mathbf{x})$ in (10) cannot be directly tested. In fact, it depends on $\boldsymbol{v} \in \mathscr{V}(\mathbf{x})$ which are unobservable. Nonetheless, the identification results derived above suggest that Θ_{I} being nonempty is a necessary condition of the single equilibrium in the data assumption in the presence of discrete common knowledge unobserved heterogeneity. More formally, under Assumptions 1 to 5, the null and alternative hypotheses in (10) lead to the following testable implications:

$$\tilde{\mathrm{H}}_0 : \Theta_{\mathrm{I}} \neq \emptyset; \quad \tilde{\mathrm{H}}_1 : \Theta_{\mathrm{I}} = \emptyset. \tag{40}$$

These testable implications are the ones that will be used to construct the statistical test described in this section.

Of course, one drawback of considering testable implications is that, for some data generating processes, $\tilde{\mathrm{H}}_0$ could be true even if H_0 is false. This limitation of the test is due to the fact that one does not observe the true $p(\mathbf{y}|\boldsymbol{v})$'s, but one knows that they belong to a set. Some of the probabilities in this set may be rationalized through a single equilibrium even if the data are generated by multiple ones.

In most cases, however, this possible drawback is not a concern. Proposition 3 states that if the number of players and/or the number of actions in the choice set are large enough compared to $|\mathcal{V}|$, testing $\tilde{\mathrm{H}}_0$ is both necessary and sufficient for H_0, except for a set of data generating processes that has Lebesgue measure zero within the space of choice probabilities $p(\mathbf{y}|\boldsymbol{v}) \; \forall \mathbf{y} \in \mathscr{Y}^N$ and $\forall \boldsymbol{v} \in \mathscr{V}$. In that sense, the testable implications are said to be generically sufficient for the single equilibrium in the data assumption.

PROPOSITION 3 (Generically sufficient testable implications). *Suppose that Assumptions 1 to 5 are satisfied. Then, if* $\sum_{n=1}^{N-1} \binom{n(J+1)}{2}\binom{J+1}{2} \geq |\mathscr{V}|$, $\Theta_I \neq \emptyset$ *is necessary and generically sufficient for* $|\mathscr{T}^*(\mathbf{x}, \boldsymbol{v})| = 1, \forall \boldsymbol{v} \in \mathscr{V}(\mathbf{x})$.

Proof. See Appendix A.4. □

A few remarks are worth making about Proposition 3 and its implications.

Remark 8 (Condition not necessary). While the condition $\sum_{n=1}^{N-1} \binom{n(J+1)}{2}\binom{J+1}{2} \geq |\mathscr{V}|$ (along with Assumptions 1–5) guarantees that $\tilde{H}_0$ is generically sufficient for H_0, it may not be necessary. In fact, the proof does not rule out that generic sufficiency could hold even if this condition is not satisfied.

Remark 9 (Small numbers of players and actions). $\sum_{n=1}^{N-1} \binom{n(J+1)}{2}\binom{J+1}{2} \geq |\mathscr{V}|$ implies that $\tilde{H}_0$ is not generically sufficient for H_0 when $N = 2$ players choose between $J + 1 = 2$ possible actions, even if $|\mathscr{V}|$ is as small as 2. However, marginally increasing the number of players or the number of actions already allows for a larger number of points in the support $\mathscr{V}$ while ensuring that $\tilde{H}_0$ is generically sufficient for H_0. For games in which three players choose between two actions, the condition for generic sufficiency is satisfied with $|\mathscr{V}| \leq 7$. With two players and three actions, up to $|\mathscr{V}| \leq 9$ may be allowed for. In both cases, the restriction $|\mathscr{V}| \leq \min\left\{(J+1)^N, |\mathscr{X}_E|\right\}$ is still satisfied provided that there are sufficient points in $|\mathscr{X}_E|$.

It is also worth emphasizing that, even in cases where the condition $\sum_{n=1}^{N-1} \binom{n(J+1)}{2}\binom{J+1}{2} \geq |\mathscr{V}|$ is not satisfied, a statistical test based on $\tilde{H}_0$ remains informative and is still of interest to applied researchers hoping to leverage multiplicity of equilibria to identify payoff functions. In fact, if one rejects $\tilde{H}_0$ then one must reject H_0. As a result, rejecting $\tilde{H}_0$ provides evidence of multiple equilibria being realized in the data that is robust to discrete common knowledge unobserved heterogeneity. In contrast, if one finds evidence of correlation between players' decisions using currently existing tests in the literature, it is not obvious whether this correlation is due to multiple equilibria or to common knowledge unobserved heterogeneity since the latter is assumed away.

Checking the testable implications of the single equilibrium in the data assumption boils down to a specification test in partial identification.[19] One can use the intersection bounds approach from Chernozhukov et al. (2013) to perform this test.[20] The details about the implementation of their procedure in the current setting are given in Appendix B.

---

[19] The idea of checking testable implications 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 (2019) for generalized regression monotonicity; or Mourifié, Henry, and Méango (2020) for the Roy model. In many cases, the proposed test is implemented via a specification test, similarly as in the current setting.

[20] A previously circulated version of the current paper used the inference method of Shi and Shum (2015) which worked well and generated simulation results qualitatively similar to the ones reported below. However, Shi and Shum's (2015) approach requires to have at least one equality condition defining the identified set. While this

5.2. Intersection Bounds Formulation

For a given $\mathbf{x}_{\text{NE}}$, the conditions defining Θ_{I} can be rewritten as inequalities that are greater or equal to zero. These inequalities depend on the joint probabilities of players' decisions $\forall \mathbf{y} \in \mathscr{Y}^N$ and $\forall \mathbf{x}_{\text{E}} \in \mathscr{X}_{\text{E}}$. Let $\mathbf{p}$ be the vector collecting all such probabilities[21], and let $\mathbf{p}^0$ be the true value of $\mathbf{p}$ in the population. For a given $\boldsymbol{\theta}$, collect the inequalities defining Θ_{I} in a $L \times 1$ vector $\mathbf{g}\left(\mathbf{p}^0, \boldsymbol{\theta}\right) \equiv \left[g_1\left(\mathbf{p}^0, \boldsymbol{\theta}\right), \ldots, g_L\left(\mathbf{p}^0, \boldsymbol{\theta}\right)\right]'$ and let $\mathscr{L} \equiv \{1, 2, \ldots, L\}$. From this notation, one can write the identified set as

$$\Theta_{\text{I}} = \left\{\boldsymbol{\theta} \in \Theta : \inf_{l \in \mathscr{L}} g_l\left(\mathbf{p}^0, \boldsymbol{\theta}\right) \geq 0\right\}. \tag{41}$$

Testing $\boldsymbol{\theta} \in \Theta_{\text{I}}$ is equivalent to testing $\inf_{l \in \mathscr{L}} g_l\left(\mathbf{p}^0, \boldsymbol{\theta}\right) \geq 0$ against $\inf_{l \in \mathscr{L}} g_l\left(\mathbf{p}^0, \boldsymbol{\theta}\right) < 0$. Chernozhukov et al. (2013) propose an estimate of the end point of a $1 - \alpha$ one-sided confidence interval for $\inf_{l \in \mathscr{L}} g_l\left(\mathbf{p}^0, \boldsymbol{\theta}\right)$, denoted $\hat{g}\left(\boldsymbol{\theta}, \alpha\right)$, such that

$$\hat{g}\left(\boldsymbol{\theta}, \alpha\right) = \inf_{l \in \mathscr{L}} \left\{g_l\left(\hat{\mathbf{p}}, \boldsymbol{\theta}\right) + \hat{c}\left(\boldsymbol{\theta}, 1 - \alpha\right)\hat{\sigma}_l\left(\boldsymbol{\theta}\right)\right\}, \tag{42}$$

where $\hat{c}\left(\boldsymbol{\theta}, 1 - \alpha\right)$ is an estimated critical value and $\hat{\sigma}_l\left(\boldsymbol{\theta}\right)$ is the estimated standard error of $g_l\left(\hat{\mathbf{p}}, \boldsymbol{\theta}\right)$. Details about $\hat{c}\left(\boldsymbol{\theta}, 1 - \alpha\right)$ and $\hat{\sigma}_l\left(\boldsymbol{\theta}\right)$ are in Appendix B. The vector $\hat{\mathbf{p}}$ is the estimate of the population probabilities $\mathbf{p}^0$ such that

$$M^{1/2}\left(\hat{\mathbf{p}} - \mathbf{p}^0\right) \overset{d}{\to} N\left(\mathbf{0}, \boldsymbol{\Sigma}\right), \tag{43}$$

where $\boldsymbol{\Sigma}$ is a block-diagonal matrix with $\mathbf{x}_{\text{E}}$-specific blocks $\boldsymbol{\Sigma}\left(\mathbf{x}_{\text{E}}\right)/\Pr\left(\mathscr{X}_{\text{E}} = \mathbf{x}_{\text{E}}\right)$. Chernozhukov et al. (2013) show that

$$\limsup_{M \to \infty} \Pr\left(\inf_{l \in \mathscr{L}} g_l\left(\mathbf{p}^0, \boldsymbol{\theta}\right) \geq \hat{g}\left(\boldsymbol{\theta}, \alpha\right)\right) \leq \alpha. \tag{44}$$

It follows that one rejects the hypothesis that $\boldsymbol{\theta} \in \Theta_{\text{I}}$ at the $1 - \alpha$ confidence level if $\hat{g}\left(\boldsymbol{\theta}, \alpha\right) < 0$.

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.[22] Let the confidence set be defined as

$$\text{CS}_M\left(\alpha\right) = \left\{\boldsymbol{\theta} \in \Theta : \hat{g}\left(\boldsymbol{\theta}, \alpha\right) \geq 0\right\}. \tag{45}$$

---

requirement is satisfied when testing (40) (i.e., testing necessary conditions for both the single equilibrium in the data assumption and the finite mixture representation of the unobserved heterogeneity), it is not met when focusing only on the conditions associated with the finite mixture representation (see Section 6.4).

[21] In practice, one should drop the probabilities associated with a given realization $\mathbf{y}$ since the probabilities over all $\mathbf{y} \in \mathscr{Y}^N$ are linearly dependent. This transformation is not made explicit in the text to alleviate notation.

[22] 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.

One can then consider the following non-randomized decision rule:

$$\xi_M = \begin{cases} 1, & \text{if } CS_M(\alpha) = \emptyset \\ 0, & \text{otherwise.} \end{cases} \tag{46}$$

Let $E_{\boldsymbol{\theta}}[\cdot]$ denote the expectation under the data generating process corresponding to $\boldsymbol{\theta}$. In particular, for $\boldsymbol{\theta} \in \Theta_I$, $E_{\boldsymbol{\theta}}[\xi_M]$ is the probability of rejecting $\tilde{H}_0 : \Theta_I \neq \emptyset$ when it is true. Proposition 4 states that the statistical test of this null hypothesis based on the decision rule ξ_M is asymptotically level α.

PROPOSITION 4 (Asymptotic level of the test). *Suppose that:*

  (i) $\mathbf{G}(\mathbf{p}, \boldsymbol{\theta}) \equiv \partial \mathbf{g}(\mathbf{p}, \boldsymbol{\theta}) / \partial \mathbf{p}'$ *is uniformly Lipschitz;*
 (ii) *the euclidean norm of each row of $\mathbf{G}(\mathbf{p}, \boldsymbol{\theta})$ is bounded away from 0 uniformly in $\mathbf{x}_{NE}$ and $\mathbf{y}$;*
(iii) *the eigenvalues of $\boldsymbol{\Sigma}$ are bounded from above and away from 0; and*
(iv) $\hat{\boldsymbol{\Sigma}}$ *is a consistent estimator of $\boldsymbol{\Sigma}$.*
     Then, for any $\boldsymbol{\theta} \in \Theta_I$, $\limsup\limits_{M \to \infty} E_{\boldsymbol{\theta}}[\xi_M] \leq \alpha$.

Proof. The result follows from Chernozhukov et al. (2013), Thm. 4, p. 694). Details are provided in Appendix A.5.   □

5.3. Computational Aspects

The conditions defining Θ_I are nonlinear in the elements of $\boldsymbol{\theta}$ and, therefore, one would typically perform a grid search over the parameter space in order to check whether $CS_M(\alpha)$ is empty. Of course, since the dimension of the parameter space increases in $|\mathscr{V}|$ – there are $|\mathscr{V}| \times (|\mathscr{V}| - 1)$ elements in $\boldsymbol{\theta}$ – the computational burden of the test increases with the number of components in the mixture over the unobserved heterogeneity. The dimension of the parameter space that one must deal with in the current test is inherited from the partial identification of the finite mixture over the unobserved heterogeneity and is therefore exactly the same as in Henry et al. (2014).

There is however a nice feature of Θ_I that can be leveraged to considerably reduce the computational burden of the test: the equality constraints in part (iii) from Proposition 2. These equality constraints considerably shrink the identified set and can be used to choose starting values from the gridded parameter space. To see this, notice that the condition rank $\{\mathbf{P}_n^\zeta(\boldsymbol{v}^l)\} = 1$ implies that all minors of order 2 constructed from $\mathbf{P}_n^\zeta(\boldsymbol{v}^l)$ must be equal to 0. Using (25), part (iii) from Proposition 2 defines quadratic expressions in reparameterizations of $\boldsymbol{\theta}$ given by $[\mathbf{e}_l - \boldsymbol{\phi}]' \boldsymbol{\Upsilon}'^{-1}$ for $l \in \{0, 1, \ldots, |\mathscr{V}| - 1\}$ that must simultaneously be equal to 0. If $|\mathscr{V}| = 2$, these equality restrictions define the roots of parabolas which can be easily computed in closed form. One can then use the values from the grid that are the closest to the roots as starting values in the grid search. If $|\mathscr{V}| = 3$, the equality restrictions define conic sections and the (reparameterized) values of $\boldsymbol{\theta}$ that belong

to Θ_I are defined by the intersections of these conic sections. If $|\mathcal{V}| > 3$ these values are at the intersections of quadric surfaces. While implementing algorithms to find the intersections of conic sections or quadric surfaces may be challenging (see e.g., Chan, 2006), the ability to initiate the grid search at values that are close to conic sections or quadric surfaces may considerably reduce the computational burden.

It is also worth describing how many inequalities must typically be checked to see if $\boldsymbol{\theta}$ belongs to $\mathrm{CS}_M(\alpha)$.[23] Parts (i) and (ii) in Proposition 2 are written in terms of convex hulls and ranges of functions as in Henry et al. (2014). Alternatively, part (i) can be written as $\min_{\mathbf{x}_E \in \mathscr{X}_E} \left\{ \gamma\left(\boldsymbol{v}^l \big| \mathbf{x}_E\right)\right\} \geq 0$ for $l = 1, \ldots, |\mathcal{V}| - 1$ and $\max_{\mathbf{x}_E \in \mathscr{X}_E} \left\{ \sum_{l=1}^{|\mathcal{V}|-1} \gamma\left(\boldsymbol{v}^l \big| \mathbf{x}_E\right)\right\} < 1$, leading to $|\mathcal{V}|$ conditions to check. Part (ii) can also be written as $\min_{\mathbf{y} \in \mathscr{Y}^N} p\left(\mathbf{y} | \boldsymbol{v}\right) \geq 0$ and $\sum_{\mathbf{y} \in \mathscr{Y}^N} p\left(\mathbf{y} | \boldsymbol{v}\right) = 1 \ \forall \boldsymbol{v} \in \mathcal{V}$. Notice, however, that the latter equality is always satisfied given the expression for $p\left(\mathbf{y} | \boldsymbol{v}\right)$ in Proposition 1 since $\sum_{\mathbf{y} \in \mathscr{Y}^N} p\left(\mathbf{y} | \mathbf{x}_E\right) = 1$. Part (ii) therefore leads to $|\mathcal{V}|$ inequalities. Finally, equalities in part (iii) of Proposition 2, more precisely the conditions that all minors of order 2 must be equal to 0, lead to two inequalities: both the minimum value of all minors of order 2 and the minimum value of the negative of all minors of order 2 must be greater than or equal to 0. Overall, there are therefore $2 \times |\mathcal{V}| + 2$ inequalities to check to see if $\boldsymbol{\theta}$ belongs to $\mathrm{CS}_M(\alpha)$.

6. MONTE CARLO SIMULATIONS

This section provides simulation evidence in order to investigate the statistical size and power of the test for the single equilibrium in the data assumption based on the testable implications. More precisely, the size of the test is the probability of rejecting $\tilde{H}_0$ (i.e., $\xi_M = 1$) when H_0 is true (i.e., there is a single equilibrium in the data for all values of $\boldsymbol{v}$). The power of the test is the probability of rejecting $\tilde{H}_0$ (i.e., $\xi_M = 1$) when H_0 is false (i.e., there are multiple equilibria in the data for some value(s) of $\boldsymbol{v}$). The robustness of the test under some levels of misspecification is also studied.

6.1. Data Generating Processes

The data generating processes are based on the simple game of market entry introduced in Example 1. The $N = 2$ version of the model as described above, as well as a $N = 3$ extension are considered. Different data generating processes are created by varying equilibrium selection mechanisms and realizations of $\mathcal{V}$. The analysis of the size and the power of the test is based on a total of 13 different cases which vary according to the number of players, the number of equilibria realized in the data, which equilibria are realized and the equilibrium selection mechanisms. Six other cases are included to study whether the test is robust

---

[23]Appendix B provides a detailed description for the case with N players, $J + 1$ actions and $|\mathcal{V}| = 2$ components in the mixture over the unobserved heterogeneity, which is implemented in the Monte Carlo simulations.

to some potential misspecifications. In particular, the simulations investigate the effect of misspecifying the number of components in the finite mixture over $\mathcal{V}$ and incorrectly discretizing the support of $\mathcal{X}_E$. The 19 different data generating processes are summarized in Table 1.

When $N = 2$, let firms' payoffs be

$$\pi_1(1, y_2, \mathbf{x}_{NE}, v_1, \varepsilon_1) = x_{NE,1} + v_1 - 4y_2 - \varepsilon_1, \tag{47}$$

$$\pi_2(1, y_1, \mathbf{x}_{NE}, v_2, \varepsilon_2) = x_{NE,2} + v_2 - 3y_1 - \varepsilon_2. \tag{48}$$

When $N = 3$, firms' payoffs are given by

$$\pi_1(1, y_2, y_3, \mathbf{x}_{NE}, v_1, \varepsilon_1) = x_{NE,1} + v_1 - 4(y_2 + y_3) - \varepsilon_1, \tag{49}$$

$$\pi_2(1, y_1, y_3, \mathbf{x}_{NE}, v_2, \varepsilon_2) = x_{NE,2} + v_2 - 3(y_1 + y_3) - \varepsilon_2, \tag{50}$$

$$\pi_3(1, y_1, y_2, \mathbf{x}_{NE}, v_3, \varepsilon_3) = x_{NE,3} + v_3 - 3.5(y_1 + y_2) - \varepsilon_3. \tag{51}$$

Similarly as before, let $p_i(1|\boldsymbol{v})$ denote the probability that $y_i = 1$, given $\mathbf{x}_{NE}$ (dropped to alleviate notation) and $\boldsymbol{v}$. Then, for $N = 2$, the BNE in pure strategies indexed by τ is such that

$$\mathbf{p}(\boldsymbol{v}, \tau) \equiv \begin{bmatrix} p_1(1|\boldsymbol{v}, \tau) \\ p_2(1|\boldsymbol{v}, \tau) \end{bmatrix} = \begin{bmatrix} \Phi(x_{NE,1} + v_1 - 4p_2(1|\boldsymbol{v}, \tau)) \\ \Phi(x_{NE,2} + v_2 - 3p_1(1|\boldsymbol{v}, \tau)) \end{bmatrix}. \tag{52}$$

For $N = 3$, it becomes

$$\mathbf{p}(\boldsymbol{v}, \tau) \equiv \begin{bmatrix} p_1(1|\boldsymbol{v}, \tau) \\ p_2(1|\boldsymbol{v}, \tau) \\ p_3(1|\boldsymbol{v}, \tau) \end{bmatrix} = \begin{bmatrix} \Phi(x_{NE,1} + v_1 - 4(p_2(1|\boldsymbol{v}, \tau) + p_3(1|\boldsymbol{v}, \tau))) \\ \Phi(x_{NE,2} + v_2 - 3(p_1(1|\boldsymbol{v}, \tau) + p_3(1|\boldsymbol{v}, \tau))) \\ \Phi(x_{NE,3} + v_3 - 3.5(p_1(1|\boldsymbol{v}, \tau) + p_2(1|\boldsymbol{v}, \tau))) \end{bmatrix}. \tag{53}$$

For most data generating processes, $\mathbf{x}_{NE} \equiv [x_{NE,1}, x_{NE,2}, x_{NE,3}]' = [1, 1, 1]'$ and $\boldsymbol{v}^l = [v_1^l, v_2^l, v_3^l]'$ for $l = 0, 1$ such that $\boldsymbol{v}^0 = [1, 0.5, 0.5]'$, $\boldsymbol{v}^1 = [1.25, 1, 1.25]'$. When studying robustness of the test, some cases add one more realization $\boldsymbol{v}^2 = [0.5, 0.25]'$ to the data generating process described above with $N = 2$. Table 2 summarizes the multiple equilibria of the model. For $N = 2$, this model admits three solutions for each $\boldsymbol{v}$. When $N = 3$, there are five solutions for each realization $\boldsymbol{v}$.

All data generating processes are such that at most two equilibria are realized in the data for each value of $\boldsymbol{v}$. With $N = 2$, these equilibria are $\mathcal{T}^*(\boldsymbol{v}^0) = \{1, 3\}$, $\mathcal{T}^*(\boldsymbol{v}^1) = \{2, 3\}$, and $\mathcal{T}^*(\boldsymbol{v}^2) = \{2, 3\}$. With $N = 3$, let $\mathcal{T}^*(\boldsymbol{v}^0) = \{4, 5\}$ and $\mathcal{T}^*(\boldsymbol{v}^1) = \{1, 2\}$. In all cases, let $\lambda(\boldsymbol{v})$ be $\lambda(\tau|\boldsymbol{v})$ for τ corresponding to the first element in $\mathcal{T}^*(\boldsymbol{v})$. Varying $\mathcal{T}^*(\boldsymbol{v})$ and $\lambda(\boldsymbol{v})$ leads to H_0 and $\tilde{H}_0$ being true or false depending on the data generating process.

When studying the size and the power of the test, there are four different realizations of $\mathbf{x}_E$. For each of these realizations, the weights associated with $\boldsymbol{v}^1$ are listed in Table 3 where $\gamma(\mathbf{x}_E) \equiv \gamma(\boldsymbol{v}^1|\mathbf{x}_E)$. Since $\gamma(\mathbf{x}_E^0) = 0.05$ and $\gamma(\mathbf{x}_E^1) = 0.95$,

TABLE 1. Description of data generating processes

Case	H_0	$\tilde{H}_0$	N	Description
				Size
S1	True	True	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{3\}; \lambda\left(\boldsymbol{v}^0\right) = 1; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{3\}; \lambda\left(\boldsymbol{v}^1\right) = 1$
S2	True	True	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{3\}; \lambda\left(\boldsymbol{v}^0\right) = 1; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2\}; \lambda\left(\boldsymbol{v}^1\right) = 1$
S3	True	True	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{1\}; \lambda\left(\boldsymbol{v}^0\right) = 1; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{3\}; \lambda\left(\boldsymbol{v}^1\right) = 1$
S4	True	True	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{1\}; \lambda\left(\boldsymbol{v}^0\right) = 1; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2\}; \lambda\left(\boldsymbol{v}^1\right) = 1$
S5	True	True	3	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{5\}; \lambda\left(\boldsymbol{v}^0\right) = 1; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2\}; \lambda\left(\boldsymbol{v}^1\right) = 1$
				Power
P1	False	False	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{1,3\}; \lambda\left(\boldsymbol{v}^0\right) = 0.25; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2,3\}; \lambda\left(\boldsymbol{v}^1\right) = 0.90$
P2	False	False	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{1,3\}; \lambda\left(\boldsymbol{v}^0\right) = 0.35; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2,3\}; \lambda\left(\boldsymbol{v}^1\right) = 0.90$
P3	False	False	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{1,3\}; \lambda\left(\boldsymbol{v}^0\right) = 0.35; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2,3\}; \lambda\left(\boldsymbol{v}^1\right) = 0.55$
P4	False	False	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{1,3\}; \lambda\left(\boldsymbol{v}^0\right) = 0.35; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2,3\}; \lambda\left(\boldsymbol{v}^1\right) = 0.05$
P5	False	False	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{1,3\}; \lambda\left(\boldsymbol{v}^0\right) = 0.85; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2,3\}; \lambda\left(\boldsymbol{v}^1\right) = 0.90$
P6	False	True	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{1,3\}; \lambda\left(\boldsymbol{v}^0\right) = 0.90; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2,3\}; \lambda\left(\boldsymbol{v}^1\right) = 0.50$
P7	False	False	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{1,3\}; \lambda\left(\boldsymbol{v}^0\right) = 0.25; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2\}; \lambda\left(\boldsymbol{v}^1\right) = 1$
P8	False	False	3	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{4,5\}; \lambda\left(\boldsymbol{v}^0\right) = 0.25; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{1,2\}; \lambda\left(\boldsymbol{v}^1\right) = 0.90$
				Robustness
R1	True	True	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{3\}; \lambda\left(\boldsymbol{v}^0\right) = 1; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{3\}; \lambda\left(\boldsymbol{v}^1\right) = 1; \mathscr{T}^*\left(\boldsymbol{v}^2\right) = \{3\}; \lambda\left(\boldsymbol{v}^2\right) = 1$
R2	False	False	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{1,3\}; \lambda\left(\boldsymbol{v}^0\right) = 0.25; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2,3\}; \lambda\left(\boldsymbol{v}^1\right) = 0.90; \mathscr{T}^*\left(\boldsymbol{v}^2\right) = \{2,3\}; \lambda\left(\boldsymbol{v}^2\right) = 0.50$
R3	True	True	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{3\}; \lambda\left(\boldsymbol{v}^0\right) = 1; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{3\}; \lambda\left(\boldsymbol{v}^1\right) = 1; \text{pool } \mathbf{x}_{NE}^2, \mathbf{x}_{NE}^3$
R4	True	True	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{3\}; \lambda\left(\boldsymbol{v}^0\right) = 1; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{3\}; \lambda\left(\boldsymbol{v}^1\right) = 1; \text{pool } \mathbf{x}_{NE}^0, \mathbf{x}_{NE}^1$
R5	False	False	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{1,3\}; \lambda\left(\boldsymbol{v}^0\right) = 0.25; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2,3\}; \lambda\left(\boldsymbol{v}^1\right) = 0.90; \text{pool } \mathbf{x}_{NE}^2, \mathbf{x}_{NE}^3$
R6	False	False	2	$\mathscr{T}^*\left(\boldsymbol{v}^0\right) = \{1,3\}; \lambda\left(\boldsymbol{v}^0\right) = 0.25; \mathscr{T}^*\left(\boldsymbol{v}^1\right) = \{2,3\}; \lambda\left(\boldsymbol{v}^1\right) = 0.90; \text{pool } \mathbf{x}_{NE}^0, \mathbf{x}_{NE}^1$

Note: $H_0 : |\mathscr{T}^*(\mathbf{x}, \boldsymbol{v})| = 1, \forall \boldsymbol{v} \in \mathscr{V}(\mathbf{x})$ and $\tilde{H}_0 : \Theta_I \neq \emptyset$. $\lambda(\boldsymbol{v})$ is $\lambda(\tau | \boldsymbol{v})$ for τ corresponding to the first element in $\mathscr{T}^*(\boldsymbol{v})$.

TABLE 2. Multiple equilibria in the model

	$N = 2$			$N = 3$	
	$\boldsymbol{\nu}^0$	$\boldsymbol{\nu}^1$	$\boldsymbol{\nu}^2$	$\boldsymbol{\nu}^0$	$\boldsymbol{\nu}^1$
$\mathbf{p}(\boldsymbol{\nu}, \tau = 1)$	(0.05, 0.91)	(0.05, 0.97)	(0.02, 0.88)	(0.03, 0.17, 0.78)	(0.01, 0.55, 0.62)
$\mathbf{p}(\boldsymbol{\nu}, \tau = 2)$	(0.50, 0.50)	(0.77, 0.38)	(0.55, 0.34)	(0.04, 0.90, 0.04)	(0.01, 0.20, 0.93)
$\mathbf{p}(\boldsymbol{\nu}, \tau = 3)$	(0.95, 0.09)	(0.90, 0.24)	(0.88, 0.08)	(0.04, 0.25, 0.68)	(0.02, 0.94, 0.13)
$\mathbf{p}(\boldsymbol{\nu}, \tau = 4)$	–	–	–	(0.49, 0.48, 0.03)	(0.75, 0.33, 0.06)
$\mathbf{p}(\boldsymbol{\nu}, \tau = 5)$	–	–	–	(0.75, 0.33, 0.06)	(0.89, 0.20, 0.06)

Note: $\mathbf{p}(\boldsymbol{\nu}, \tau)$ is the vector containing each player's probability of choosing $y_i = 1$ given $\mathbf{x}_{\text{NE}}$ (dropped to alleviate notation) and $\boldsymbol{\nu}$ in equilibrium τ.

TABLE 3. Mixture weights for unobserved heterogeneity

	$\mathbf{x}_{\text{E}}^0$	$\mathbf{x}_{\text{E}}^1$	$\mathbf{x}_{\text{E}}^2$	$\mathbf{x}_{\text{E}}^3$
$\gamma(\mathbf{x}_{\text{E}})$	0.05	0.95	0.50	0.65

Note: $\gamma(\mathbf{x}_{\text{E}}) \equiv \gamma(\boldsymbol{\nu}^1 | \mathbf{x}_{\text{E}})$.

it follows that $\phi^0 = 0.05$ and $\Upsilon^0 = 0.90$. These weights will be modified for some of the robustness analysis provided below.

For a given Monte Carlo sample, the joint conditional choice probabilities in $\mathbf{p}^0$ are estimated using a simple frequency count estimator. Since estimated choice probabilities equal to 0 may lead to ill-defined test statistics, 0's are replaced with 10^{-6} and the vector of choice probabilities given $\mathbf{x}_{\text{E}}$ is normalized to sum to 1. Moreover, since the sets $\widetilde{\mathcal{X}}_{\text{E}}$ and $\widetilde{\mathcal{Y}^N}$ are arbitrary, let $\tilde{\mathbf{x}}_{\text{E}}^0 = \mathbf{x}_{\text{E}}^0$, $\tilde{\mathbf{x}}_{\text{E}}^1 = \mathbf{x}_{\text{E}}^1$, and let $\tilde{\mathbf{y}}^1$ be the value of $\mathbf{y} \in \mathcal{Y}^N$ that maximizes $\hat{p}\left(\mathbf{y} | \tilde{\mathbf{x}}_{\text{E}}^1\right) - \hat{p}\left(\mathbf{y} | \tilde{\mathbf{x}}_{\text{E}}^0\right)$. This is done to avoid $\hat{p}\left(\tilde{\mathbf{y}}^1 | \tilde{\mathbf{x}}_{\text{E}}^1\right) - \hat{p}\left(\tilde{\mathbf{y}}^1 | \tilde{\mathbf{x}}_{\text{E}}^0\right) = 0$ which would cause the procedure to fail.

In order to compute ξ_M, one must span the parameter space Θ, at least until $\text{CS}_M(\alpha)$ is found to be nonempty. Simulations below use the grid $\{0.005, 0.01, \ldots, 0.99, 0.995\}$ for both ϕ and Υ. As described in Section 5.3, equality restrictions from Proposition 2 part (iii) are used to choose starting values on the grid. If these starting values are not in $\text{CS}_M(\alpha)$, the method considers each possible value of $\boldsymbol{\theta}$ in this discretized parameter space (subject to $\phi + \Upsilon < 1$) and computes $\hat{g}(\boldsymbol{\theta}, \alpha)$ as in (42) where the details about the construction of the critical values and the standard errors are in Appendix B. When computing the critical values, $R = 500$ draws are used. The results report different sample sizes for each value of $\mathbf{x}_{\text{E}}$ denoted by $M(\mathbf{x}_{\text{E}}) \in \{50, 250, 500, 1{,}000\}$. In most cases, let $|\mathcal{X}_{\text{E}}| = 4$ (such that $M \in \{200, 1{,}000, 2{,}000, 4{,}000\}$). The only exceptions are cases R3–R6 in which the incorrectly discretized support of $\mathcal{X}_{\text{E}}$ is such that $|\mathcal{X}_{\text{E}}| = 3$ (and $M \in \{150, 750, 1{,}500, 3{,}000\}$). The rejection probabilities of the test, for $\alpha = 0.1$ and $\alpha = 0.05$, are computed using 250 Monte Carlo samples.

TABLE 4. Size of the test

Case	$M(\mathbf{x}_E)=50$		$M(\mathbf{x}_E)=250$		$M(\mathbf{x}_E)=500$		$M(\mathbf{x}_E)=1{,}000$	
	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.1$	$\alpha=0.05$
All restrictions								
S1	17.6	13.6	20.8	18.4	9.6	7.2	3.6	2.0
S2	54.4	47.2	15.2	14.4	4.0	4.0	2.4	1.6
S3	24.4	24.4	0.0	0.0	0.0	0.0	0.0	0.0
S4	29.2	28.4	0.0	0.0	0.0	0.0	0.0	0.0
S5	3.2	2.0	30.4	23.2	29.6	23.6	14.4	10.4
Finite mixture restrictions only								
S1	6.8	2.8	12.4	10.4	6.8	5.6	0.8	0.8
S2	50.0	41.6	12.8	12.8	2.8	2.8	0.0	0.0
S3	24.4	24.4	0.0	0.0	0.0	0.0	0.0	0.0
S4	28.4	28.4	0.0	0.0	0.0	0.0	0.0	0.0
S5	1.6	1.6	16.4	15.6	13.6	12.8	1.6	1.2

Notes: Rejection probabilities in % computed over 250 Monte Carlo samples. Critical values to construct the confidence set use 500 draws. The finite mixture restrictions hold in all data generating processes.

6.2. Size of the Test

Cases S1–S5 are such that the data are generated from a single equilibrium at $\boldsymbol{\nu}^0$ and $\boldsymbol{\nu}^1$. For these cases, $\Theta_I \neq \emptyset$. The probabilities of rejecting $\tilde{H}_0$ are reported in the first part of Table 4. A few comments are worth making. First, cases S1–S4 confirm that the test defined by ξ_M in (46) is asymptotically level α, as expected from Proposition 4. However, the test is somewhat conservative. This observation is in fact a common feature of specification tests in partial identification that are 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.

Second, the main difference between case S5 and the previous ones is that the former is a game between $N = 3$ players instead of 2. Since there are $2^N - 1$ probabilities to be estimated for each $\mathbf{x}_E$, the case with $N = 3$ requires greater sample sizes for the asymptotic properties to apply. Notice that the rejection probabilities when $M(\mathbf{x}_E) = 50$ are much smaller than the ones obtained for larger sample sizes. However, this observation seems to be due to several estimated probabilities in $\hat{\mathbf{p}}$ being zeros (therefore set to 10^{-6} before normalizing the sum to 1) in small samples. The results for $M(\mathbf{x}_E) = 50$ should therefore be interpreted with caution. The test does over reject in case S5 even at $M(\mathbf{x}_E) = 1{,}000$, but the probability of rejecting typically decreases as sample size increases.

Testing whether Θ_I is empty amounts to checking the testable implications associated with both the single equilibrium in the data assumption and the finite

mixture representation of the unobserved heterogeneity. One may therefore wonder if rejecting $\tilde{H}_0$ is due to rejecting the single equilibrium in the data assumption or the finite and discrete unobserved heterogeneity assumptions. An appealing feature of the proposed test is that it allows the researcher to separately check whether the restrictions due to the finite mixture representation are satisfied even if one rejects $\Theta_I \neq \emptyset$.

The second half of Table 4 reports the probability of rejecting the finite mixture assumptions, which are satisfied in all data generating processes. These probabilities are computed over the same Monte Carlo samples as when checking all restrictions. The only difference is that $\hat{g}(\theta, \alpha)$ is now constructed using only the restrictions associated with the finite mixture over $\mathcal{V}$. The results confirm that, except in a few cases, a non negligible fraction of the rejections of $\tilde{H}_0$ when H_0 is true are due to the testable implications of the single equilibrium assumption being violated.

6.3. Power of the Test

Cases P1–P8 are based on the same data generating processes as for cases S1–S5, but using different equilibrium selection mechanisms. All cases except P7 mix two equilibria for both $\boldsymbol{v}^0$ and $\boldsymbol{v}^1$. Case P7 is such that there are two equilibria realized in the data for $\boldsymbol{v}^0$, but a single one for $\boldsymbol{v}^1$. All eight cases violate the single equilibrium in the data assumption and are such that H_0 is false. For all cases except P6, $\tilde{H}_0$ is also false. In fact, for games between two players choosing one of two actions with two components in the mixture over $\mathcal{V}$, Proposition 3 implies that the set of data generating processes for which $\tilde{H}_0$ is true even if H_0 is false does not have Lebesgue measure zero. Case P6 is included as an example of such a data generating process.

While the test has power approaching 1 in large samples for most data generating processes reported in the table, the power is actually quite low in cases P5 and P6 even for large sample sizes. The power issue associated with P6 is as expected: the data can be rationalized by a single equilibrium ($\tilde{H}_0$ is true) even if the data are generated from multiple equilibria (H_0 is false). In this case, one expects the test to reject $\tilde{H}_0$ with probability at most α asymptotically since the data are still rationalizable by a single equilibrium.

The rejection probabilities associated with case P5 suggest that the test may have much better power for some alternative equilibrium selection mechanisms than others. Further investigation of the properties of this data generating process suggest that the low power associated with P5 is due to the fact that, even if $\Theta_I = \emptyset$, the data are "almost" rationalizable by a single equilibrium not too far from the true equilibrium choice probabilities. The intuition can be understood via Figure 2. The singleton satisfying conditions (iii) in Proposition 2 falls outside of, but very close to the shaded area defined by conditions (i) and (ii). It is therefore not surprising to find that the test has low power in this case.

Similarly as when analyzing the size of the test, the second panel of Table 5 is used to assess whether the test successfully rejects $\tilde{H}_0$ based on a violation

TABLE 5. Power of the test

	$M(\mathbf{x}_E) = 50$		$M(\mathbf{x}_E) = 250$		$M(\mathbf{x}_E) = 500$		$M(\mathbf{x}_E) = 1,000$	
Case	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$
	All restrictions							
P1	86.4	82.4	100.0	100.0	100.0	100.0	100.0	100.0
P2	84.6	77.6	100.0	100.0	100.0	100.0	100.0	100.0
P3	71.6	63.2	100.0	99.6	100.0	100.0	100.0	100.0
P4	60.8	55.6	94.0	90.0	99.6	98.8	100.0	100.0
P5	6.4	4.8	4.0	2.4	7.6	3.2	20.0	11.2
P6	6.4	5.6	0.0	0.0	0.0	0.0	0.0	0.0
P7	88.4	83.6	100.0	100.0	100.0	100.0	100.0	100.0
P8	15.2	11.6	62.8	52.4	88.8	75.6	98.8	97.6
	Finite mixture restrictions only							
P1	3.2	3.2	0.0	0.0	0.0	0.0	0.0	0.0
P2	2.8	2.8	0.0	0.0	0.0	0.0	0.0	0.0
P3	4.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0
P4	18.4	18.4	0.0	0.0	0.0	0.0	0.0	0.0
P5	4.4	4.4	0.0	0.0	0.0	0.0	0.0	0.0
P6	5.6	5.6	0.0	0.0	0.0	0.0	0.0	0.0
P7	2.8	2.8	0.0	0.0	0.0	0.0	0.0	0.0
P8	9.2	9.2	10.8	9.2	5.2	2.8	1.6	0.8

Notes: Rejection probabilities in % computed over 250 Monte Carlo samples. Critical values to construct the confidence set use 500 draws. The finite mixture restrictions hold in all data generating processes.

of the single equilibrium in the data restrictions or the failure to rationalize the data via a finite mixture over the unobserved heterogeneity. The finite mixture restrictions hold in the data generating processes and these restrictions should be seldom rejected in large samples. The results show that the test is successful at separating the two types of conditions. In large samples, the test almost always rejects the testable implications of the single equilibrium in the data assumption, but fails to reject the testable implications of the finite mixture representation.

6.4. Robustness

Finally, because the proposed test is based on a specification test, it is worth exploring the approach's sensitivity to some potential misspecifications. In particular, simulations reported below provide evidence that Θ_I is not de facto empty as soon as some of the assumptions maintained about the data generating process are not exactly satisfied. The following experiences explore some level of misspecification with respect to the support of the unobserved heterogeneity and the variable satisfying the exclusion restriction.

TABLE 6. Mixture weights for unobserved heterogeneity (robustness)

	$\mathbf{x}_E^0$	$\mathbf{x}_E^1$	$\mathbf{x}_E^2$	$\mathbf{x}_E^3$
$\gamma\left(\boldsymbol{\nu}^1 \middle\vert \mathbf{x}_E\right)$	0.05	0.85	0.25	0.45
$\gamma\left(\boldsymbol{\nu}^2 \middle\vert \mathbf{x}_E\right)$	0.05	0.10	0.25	0.20

So far, the number of components in the finite mixture over the unobserved heterogeneity, i.e., $|\mathcal{V}|$, has been assumed to be known.[24] Cases R1 and R2 investigate the effect of incorrectly assuming $|\mathcal{V}| = 2$ when $|\mathcal{V}| = 3$ in the data generating process (with mixture weights in Table 6). While in R1, the data are generated from a single equilibrium for each $\boldsymbol{\nu}$, there are two equilibria for each $\boldsymbol{\nu}$ in R2.

The rejection probabilities computed via Monte Carlo simulations for cases R1 and R2 are reported in the corresponding rows of Table 7. This type of misspecification does not alter the properties of the test. In R1, even if $|\mathcal{V}| = 3$ in the data generating process, $\Theta_I \neq \emptyset$ which suggests that the data are still rationalizable by a single equilibrium and a finite mixture counting two components. In that case, the misspecified test still achieves asymptotic size control. In R2, where there are multiple equilibria in the data, the data cannot be rationalized by a single equilibrium and a finite mixture over two components of unobserved heterogeneity, i.e., $\Theta_I = \emptyset$. Once again, the power of the misspecified test tends to 1 as the sample size increases.

Interestingly, for both R1 and R2, one rarely rejects the testable restrictions associated with $|\mathcal{V}| = 2$ even if the true mixture is such that $|\mathcal{V}| = 3$, at least in large samples. In other words, the data are still rationalizable by a finite mixture over two components even if there are three components in the generating process. However, it is easy to see that the current test would not be robust to incorrectly assuming $|\mathcal{V}| = 2$ when in fact $|\mathcal{V}| = 1$. Asymptotically, $p(\mathbf{y}|\mathbf{x}_E)$ would not vary with $\mathbf{x}_E$ which implies that $Q(\mathbf{x}_E)$, $L_0(\mathbf{y})$ and $L_1(\mathbf{y})$ are not defined and $\Theta_I = \emptyset$ regardless of the number of equilibria realized in the data.

Cases R3–R6 evaluate the effect of incorrectly discretizing $\mathscr{X}_E$.[25] This experiment is meant to address the concern that some variables used as proxy for the unobserved heterogeneity may be continuous in nature and discretized by the applied researcher. In order to assess the effect of an incorrect discretization of the support of such proxy variables in a simple way, suppose that some values of $\mathbf{x}_E$ are pooled together when estimating $\mathbf{p}^0$. Cases R3 and R5 pool the true $\mathbf{x}_E^2$ and $\mathbf{x}_E^3$; R4 and R6 pool the true $\mathbf{x}_E^0$ and $\mathbf{x}_E^1$. In all cases, the test is performed after

---

[24]As mentioned in Lemma 1, this number is point-identified from the rank of a matrix. Although one could perform a rank test to check whether the maintained number of components is rejected by the sample, the current procedure is not augmented with such a pre-testing approach in order to avoid issues related with testing multiple null hypotheses.

[25]While properly discretizing $\mathscr{X}_{NE}$ is also an important decision to be made, such potential misspecification is not included in the simulations. Pooling together values of $\mathbf{x}_{NE}$ associated with significantly different equilibria would likely be problematic not only for a test of the single equilibrium in the data assumption, but also for the estimation of the primitives of the model.

TABLE 7. Robustness of the test

Case	$M(\mathbf{x}_E) = 50$		$M(\mathbf{x}_E) = 250$		$M(\mathbf{x}_E) = 500$		$M(\mathbf{x}_E) = 1{,}000$	
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$
All restrictions								
R1	10.0	4.4	23.2	20.8	10.0	7.6	7.6	6.4
R2	64.4	56.8	100.0	100.0	100.0	100.0	100.0	100.0
R3	18.4	13.6	20.4	18.8	10.0	6.8	3.2	2.0
R4	25.2	22.4	5.2	4.4	4.4	3.2	2.4	1.6
R5	86.0	82.0	100.0	100.0	100.0	100.0	100.0	100.0
R6	8.0	5.6	10.4	7.2	9.2	6.8	10.4	6.4
Finite mixture restrictions only								
R1	7.2	2.8	11.2	6.0	4.8	3.6	0.4	0.4
R2	2.8	2.8	0.0	0.0	0.0	0.0	0.0	0.0
R3	6.8	4.4	13.2	10.0	6.8	6.0	0.8	0.8
R4	5.6	3.2	0.8	0.8	0.0	0.0	0.0	0.0
R5	3.2	3.2	0.0	0.0	0.0	0.0	0.0	0.0
R6	0.4	0.4	0.4	0.4	0.0	0.0	0.0	0.0

Notes: Rejection probabilities in % computed over 250 Monte Carlo samples. Critical values to construct the confidence set use 500 draws. The finite mixture restrictions hold in all data generating processes.

having estimated the joint distributions conditional on the remaining three values. The data generating processes for R3 and R4 are exactly the same as in S1 (i.e., both H_0 and $\tilde{H}_0$ are true). For R5 and R6, the data are generated as in P1 (i.e., both H_0 and $\tilde{H}_0$ are false).

The rejection probabilities are also reported in Table 7. While incorrectly pooling $\mathbf{x}_E^2$ and $\mathbf{x}_E^3$ does not affect much the size and the power of the test, the wrong discretization that fails to separate $\mathbf{x}_E^0$ and $\mathbf{x}_E^1$ can lead to important changes in the rejection probabilities. In fact, the rejection probabilities for R3 and R5 are very similar to the ones in their properly discretized counterparts (respectively S1 and P1). However, this similarity does not hold for R4 and R6. This important distinction across the alternative discretizations may be due to the fact that the differences between $\gamma(\mathbf{x}_E)$'s is much larger for $\mathbf{x}_E^0$ vs. $\mathbf{x}_E^1$ than for $\mathbf{x}_E^2$ vs. $\mathbf{x}_E^3$.

Finally, in cases R1, R4, and R6, the procedure fails in some rare instances (at most three out of 250 Monte Carlo samples, almost exclusively when $M(\mathbf{x}_E) = 50$). Such failures were not encountered in cases S1–S5 nor P1–P8.

7. CONCLUDING REMARKS

To sum up, the test of the single equilibrium in the data assumption presented in the current paper addresses two important issues associated with the procedures previously proposed in the literature. First, it allows for common knowledge

payoff-relevant unobservables, which is an extra source of plausible correlation between players' decisions beyond multiple equilibria being realized in the data. Second, the proposed test 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 useful for empirical researchers interested in testing degenerate equilibrium selection in hope of leveraging multiple equilibria as a source of variation to identify payoffs 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 (which implies restrictions that are separately testable). 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 testable implications are generically sufficient for the single equilibrium in the data assumption under some verifiable conditions. The test boils down to a specification test and it can be implemented using the intersection bounds framework of Chernozhukov et al. (2013) which has nice properties corroborated in simulations.

APPENDIX A. PROOFS

A.1. Proof of Lemma 1

While Henry et al. (2014) only consider mixtures of marginal distributions, their results also apply to mixtures of joint distributions. Notice that the (i,j)th element of $\Delta \mathbf{P}(\mathbf{x}_{NE})$ can be written as

$$\sum_{l=0}^{|\mathscr{V}(\mathbf{x}_{NE})|-1} p\left(\mathbf{y}^j \Big| \mathbf{x}_{NE}, \boldsymbol{v}^l\right) \left[\gamma\left(\boldsymbol{v}^l \Big| \mathbf{x}_{NE}, \mathbf{x}_E^i\right) - \gamma\left(\boldsymbol{v}^l \Big| \mathbf{x}_{NE}, \mathbf{x}_E^0\right)\right]. \tag{A.1}$$

Moreover, since $\gamma\left(\boldsymbol{v}^0 \Big| \mathbf{x}_{NE}, \mathbf{x}_E\right) = 1 - \sum_{l=1}^{|\mathscr{V}(\mathbf{x}_{NE})|-1} \gamma\left(\boldsymbol{v}^l \Big| \mathbf{x}_{NE}, \mathbf{x}_E\right) \ \forall \mathbf{x}_E \in \mathscr{X}_E$, (A.1) is equivalent to

$$\sum_{l=1}^{|\mathscr{V}(\mathbf{x}_{NE})|-1} p\left(\mathbf{y}^j \Big| \mathbf{x}_{NE}, \boldsymbol{v}^l\right) \left[\gamma\left(\boldsymbol{v}^l \Big| \mathbf{x}_{NE}, \mathbf{x}_E^i\right) - \gamma\left(\boldsymbol{v}^l \Big| \mathbf{x}_{NE}, \mathbf{x}_E^0\right)\right]$$
$$- p\left(\mathbf{y}^j \Big| \mathbf{x}_{NE}, \boldsymbol{v}^0\right) \sum_{l=1}^{|\mathscr{V}(\mathbf{x}_{NE})|-1} \left[\gamma\left(\boldsymbol{v}^l \Big| \mathbf{x}_{NE}, \mathbf{x}_E^i\right) - \gamma\left(\boldsymbol{v}^l \Big| \mathbf{x}_{NE}, \mathbf{x}_E^0\right)\right]. \tag{A.2}$$

In other words, the matrix $\Delta \mathbf{P}(\mathbf{x}_{NE})$ can be written as the product of two matrices. The first one is a $(|\mathscr{X}_E| - 1) \times (|\mathscr{V}(\mathbf{x}_{NE})| - 1)$ matrix with (i,j)th element $\gamma\left(\boldsymbol{v}^j \Big| \mathbf{x}_{NE}, \mathbf{x}_E^i\right) - \gamma\left(\boldsymbol{v}^j \Big| \mathbf{x}_{NE}, \mathbf{x}_E^0\right)$. The second one is a $(|\mathscr{V}(\mathbf{x}_{NE})| - 1) \times (J+1)^N$ matrix with (i,j)th element $p\left(\mathbf{y}^j \Big| \mathbf{x}_{NE}, \boldsymbol{v}^i\right) - p\left(\mathbf{y}^j \Big| \mathbf{x}_{NE}, \boldsymbol{v}^0\right)$. By Assumption 5(vi), the second matrix is full row rank. The rank of the product of the two matrices is therefore equal to the rank of the first matrix, which is $|\mathscr{V}(\mathbf{x}_{NE})| - 1$ by Assumption 5(v).

A.2. Proof of Proposition 1

For any $\mathbf{y}^{\zeta} \in \mathscr{Y}^{N^{\zeta}}$, $\mathscr{N}^{\zeta} \subseteq \mathscr{N}$, $\mathbf{x}_{\mathrm{E}} \in \mathscr{X}_{\mathrm{E}}$, one can write

$$p\left(\mathbf{y}^{\zeta} \middle| \mathbf{x}_{\mathrm{E}}\right) = p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{0}\right) + \sum_{l=1}^{|\mathscr{V}|-1} \left[p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{l}\right) - p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{0}\right)\right] \gamma\left(\mathbf{v}^{l} \middle| \mathbf{x}_{\mathrm{E}}\right). \tag{A.3}$$

Evaluating (A.3) at $\mathbf{x}_{\mathrm{E}}^{0}$, one gets

$$p\left(\mathbf{y}^{\zeta} \middle| \mathbf{x}_{\mathrm{E}}^{0}\right) = p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{0}\right) + \sum_{l=1}^{|\mathscr{V}|-1} \left[p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{l}\right) - p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{0}\right)\right] \gamma\left(\mathbf{v}^{l} \middle| \mathbf{x}_{\mathrm{E}}^{0}\right). \tag{A.4}$$

First, consider the expression of $p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{l}\right)$. Some more notation is needed. Let $\boldsymbol{\delta}\left(\mathbf{y}^{\zeta}\right)$ be the $(|\mathscr{V}|-1) \times 1$ vector with lth element $p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{l}\right) - p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{0}\right)$, and let $\boldsymbol{\Upsilon}^{c}\left(\tilde{\mathbf{x}}_{\mathrm{E}}\right)$ be the column of $\boldsymbol{\Upsilon}$ corresponding to $\tilde{\mathbf{x}}_{\mathrm{E}}$. Evaluating (A.3) at any $\tilde{\mathbf{x}}_{\mathrm{E}} \in \tilde{\mathscr{X}}_{\mathrm{E}}$ and subtracting (A.4) from it, one can write

$$p\left(\mathbf{y}^{\zeta} \middle| \tilde{\mathbf{x}}_{\mathrm{E}}\right) - p\left(\mathbf{y}^{\zeta} \middle| \mathbf{x}_{\mathrm{E}}^{0}\right) = \boldsymbol{\Upsilon}^{c}\left(\tilde{\mathbf{x}}_{\mathrm{E}}\right)' \boldsymbol{\delta}\left(\mathbf{y}^{\zeta}\right) \tag{A.5}$$

such that the $\mathbf{y}^{\zeta}$-specific column of $\boldsymbol{\Delta}\tilde{\mathbf{P}}^{\zeta}$ is $\boldsymbol{\Delta}\tilde{\mathbf{p}}^{\zeta,c}\left(\mathbf{y}^{\zeta}\right) = \boldsymbol{\Upsilon}'\boldsymbol{\delta}\left(\mathbf{y}^{\zeta}\right)$. Since $\boldsymbol{\Upsilon}$ is invertible, it follows that $\boldsymbol{\delta}\left(\mathbf{y}^{\zeta}\right) = \boldsymbol{\Upsilon}'^{-1}\boldsymbol{\Delta}\tilde{\mathbf{p}}^{\zeta,c}\left(\mathbf{y}^{\zeta}\right)$. Rearranging (A.4), $p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{0}\right)$ can be written as

$$p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{0}\right) = p\left(\mathbf{y}^{\zeta} \middle| \mathbf{x}_{\mathrm{E}}^{0}\right) - \boldsymbol{\phi}'\boldsymbol{\Upsilon}'^{-1}\boldsymbol{\Delta}\tilde{\mathbf{p}}^{\zeta,c}\left(\mathbf{y}^{\zeta}\right). \tag{A.6}$$

Furthermore, for any $\mathbf{v}^{l}$ with $l \neq 0$, $p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{l}\right) = p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{0}\right) + p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{l}\right) - p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{0}\right)$ such that

$$p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{l}\right) = p\left(\mathbf{y}^{\zeta} \middle| \mathbf{x}_{\mathrm{E}}^{0}\right) - \boldsymbol{\phi}'\boldsymbol{\Upsilon}'^{-1}\boldsymbol{\Delta}\tilde{\mathbf{p}}^{\zeta,c}\left(\mathbf{y}^{\zeta}\right) + \mathbf{e}_{l}'\boldsymbol{\Upsilon}'^{-1}\boldsymbol{\Delta}\tilde{\mathbf{p}}^{\zeta,c}\left(\mathbf{y}^{\zeta}\right). \tag{A.7}$$

It follows that $p\left(\mathbf{y}| \mathbf{v}^{l}\right)$ for any $l = 0, 1, \ldots, |\mathscr{V}|-1$ is written as

$$p\left(\mathbf{y}^{\zeta} \middle| \mathbf{v}^{l}\right) = p\left(\mathbf{y}^{\zeta} \middle| \mathbf{x}_{\mathrm{E}}^{0}\right) + [\mathbf{e}_{l} - \boldsymbol{\phi}]'\,\boldsymbol{\Upsilon}'^{-1}\boldsymbol{\Delta}\tilde{\mathbf{p}}^{\zeta,c}\left(\mathbf{y}^{\zeta}\right) \tag{A.8}$$

which is the expression in Proposition 1.

Second, consider the expression for $\boldsymbol{\gamma}\left(\mathbf{x}_{\mathrm{E}}\right)$. Since $\boldsymbol{\gamma}\left(\mathbf{x}_{\mathrm{E}}\right)$ does not depend on $\mathbf{y}$, the expression can be derived for any subset $\mathscr{N}^{\zeta} \subseteq \mathscr{N}$. Without loss of generality, consider $\mathscr{N}$. Using (A.8), the $(|\mathscr{V}|-1) \times (J+1)^{N}$ matrix with (i,j)th element $p\left(\mathbf{y}^{j} \middle| \mathbf{v}^{i}\right) - p\left(\mathbf{y}^{j} \middle| \mathbf{v}^{0}\right)$ can be written as $\boldsymbol{\Upsilon}'^{-1}\boldsymbol{\Delta}\tilde{\mathbf{P}}$. Therefore, subtracting (A.4) from (A.3) for all $\mathbf{y} \in \mathscr{Y}^{N}$, the $\mathbf{x}_{\mathrm{E}}$-specific row of $\boldsymbol{\Delta}\tilde{\mathbf{P}}$ is $\boldsymbol{\Delta}\tilde{\mathbf{p}}^{r}\left(\mathbf{x}_{\mathrm{E}}\right)' = [\boldsymbol{\gamma}\left(\mathbf{x}_{\mathrm{E}}\right) - \boldsymbol{\phi}]'\,\boldsymbol{\Upsilon}'^{-1}\boldsymbol{\Delta}\tilde{\mathbf{P}}$. Since $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Delta}\tilde{\mathbf{P}}$ are both invertible, it follows that

$$\boldsymbol{\gamma}\left(\mathbf{x}_{\mathrm{E}}\right) = \boldsymbol{\phi} + \boldsymbol{\Upsilon}\left[\boldsymbol{\Delta}\tilde{\mathbf{P}}'\right]^{-1}\boldsymbol{\Delta}\tilde{\mathbf{p}}^{r}\left(\mathbf{x}_{\mathrm{E}}\right) \tag{A.9}$$

which is the expression in Proposition 1.

A.3. Proof of Proposition 2

Parts (i) and (ii) are direct applications of Henry et al. (2014), Sects. 3 and 4.2). They ensure that the probabilities defining the mixture over $\mathcal{V}$ as written in Proposition 1 are well defined. Part (i) follows from $\gamma\left(\boldsymbol{v}^l \big| \mathbf{x}_E\right) \geq 0$ for $l = 1, \ldots, |\mathcal{V}| - 1$ and $\sum_{l=1}^{|\mathcal{V}|-1} \gamma\left(\boldsymbol{v}^l \big| \mathbf{x}_E\right) < 1 \; \forall \mathbf{x}_E \in \mathcal{X}_E$. Part (ii) is equivalent to $0 \leq p\left(\mathbf{y} | \boldsymbol{v}\right) \leq 1 \; \forall \mathbf{y} \in \mathcal{Y}^N$ and $\forall \boldsymbol{v} \in \mathcal{V}$.

Part (iii) is an application of Kasahara and Shimotsu (2014, Prop. 4). It states that there is only one component in the finite mixture over τ and, therefore, players' decisions are independent given $\mathbf{x}_{NE}$ and $\boldsymbol{v}$. In fact, since players' decisions are conditionally independent, the matrix $\mathbf{P}_n^\zeta(\boldsymbol{v})$ can be written as the product of two rank-1 matrices: the column vector collecting $p\left(\mathbf{y}_n^\zeta \big| \boldsymbol{v}\right) \; \forall \mathbf{y}_n^\zeta \in \mathcal{Y}^n$ and the row vector collecting $p\left(y_{n+1} \big| \boldsymbol{v}\right) \; \forall y_{n+1} \in \mathcal{Y}$. As a result, rank $\left\{\mathbf{P}_n^\zeta(\boldsymbol{v})\right\} = 1$. Following Dawid (1979, p. 5), one can define the joint independence of N random variables inductively. In the current setting, players' decisions are independent if player 2's decisions are independent from those of player 1, player 3's decisions are independent from those of players 1 and 2, etc. This is the reason why, for a given $\boldsymbol{v}$, it suffices to consider the ranks of matrices $\mathbf{P}_n^\zeta(\boldsymbol{v})$ for $n = 1, \ldots, N-1$. By writing the elements of $\mathbf{P}_n^\zeta(\boldsymbol{v})$ as in (25), this condition defines restrictions on $\boldsymbol{\phi}, \boldsymbol{\Upsilon}$ in terms of point identified probabilities.

A.4. Proof of Proposition 3

Necessity simply follows from the definition of Θ_I. Let $\boldsymbol{\theta}^0$ be the value of $\boldsymbol{\theta}$ that corresponds to the true data generating process with choice probabilities denoted $p^0(\mathbf{y} | \boldsymbol{v}) \; \forall \mathbf{y} \in \mathcal{Y}^N$ and $\forall \boldsymbol{v} \in \mathcal{V}$. If $\left|\mathcal{T}^*(\mathbf{x}, \boldsymbol{v})\right| = 1, \forall \boldsymbol{v} \in \mathcal{V}(\mathbf{x})$ is true, then $\boldsymbol{\theta}^0 \in \Theta_I$ and, as a result, $\Theta_I \neq \emptyset$ is true.

To show that $\Theta_I \neq \emptyset$ being true is generically sufficient for $\left|\mathcal{T}^*(\mathbf{x}, \boldsymbol{v})\right| = 1, \forall \boldsymbol{v} \in \mathcal{V}(\mathbf{x})$, one can proceed as follows. First, for some arbitrary $\bar{\boldsymbol{\theta}} \neq \boldsymbol{\theta}^0$ one can write the corresponding choice probabilities denoted $\bar{p}\left(\mathbf{y}^\zeta \big| \boldsymbol{v}\right) \; \forall \mathbf{y}^\zeta \in \mathcal{Y}^{N^\zeta}, \forall \mathcal{N}^\zeta \subseteq \mathcal{N}, \forall \boldsymbol{v} \in \mathcal{V}$ as functions of $\bar{\boldsymbol{\theta}}$, $\boldsymbol{\theta}^0$ and the true choice probabilities. Then, if $\Theta_I \neq \emptyset$ fails to be sufficient for $\left|\mathcal{T}^*(\mathbf{x}, \boldsymbol{v})\right| = 1 \; \forall \boldsymbol{v} \in \mathcal{V}(\mathbf{x})$, it must be the case that $\exists \bar{\boldsymbol{\theta}}$ that belongs to Θ_I when $\boldsymbol{\theta}^0$ does not. To show generic sufficiency, one can therefore show that the set of true data generating processes for which such $\bar{\boldsymbol{\theta}}$ exists can be characterized via some algebraic conditions and therefore has Lebesgue measure zero within the space of choice probabilities $p(\mathbf{y} | \boldsymbol{v}) \; \forall \mathbf{y} \in \mathcal{Y}^N$ and $\forall \boldsymbol{v} \in \mathcal{V}$. Somewhat similar algebraic arguments, although applied to a different context, can be found in Allman, Mathias, and Rhodes (2009).

In order to write $\bar{p}\left(\mathbf{y}^\zeta \big| \boldsymbol{v}\right)$ as functions of $\bar{\boldsymbol{\theta}}, \boldsymbol{\theta}^0$ and the true choice probabilities, evaluate (22) in Proposition 1 at $\bar{\boldsymbol{\theta}}$ to get

$$\bar{p}\left(\mathbf{y}^\zeta \big| \boldsymbol{v}^l\right) = p\left(\mathbf{y}^\zeta \big| \mathbf{x}_E^0\right) + \left[\mathbf{e}_l - \bar{\boldsymbol{\phi}}\right]' \bar{\boldsymbol{\Upsilon}}'^{-1} \Delta \tilde{\mathbf{p}}^{\zeta, c}\left(\mathbf{y}^\zeta\right), \tag{A.10}$$

where $p\left(\mathbf{y}^\zeta \big| \mathbf{x}_E^0\right)$ and $\Delta \tilde{\mathbf{p}}^{\zeta, c}\left(\mathbf{y}^\zeta\right)$ are functions of $\boldsymbol{\theta}^0$ and the true $p^0\left(\mathbf{y}^\zeta \big| \boldsymbol{v}\right)$ for $\boldsymbol{v} \in \mathcal{V}$. In fact, one can write

$$p\left(\mathbf{y}^\zeta \big| \mathbf{x}_E^0\right) = p^0\left(\mathbf{y}^\zeta \big| \boldsymbol{v}^0\right) + \boldsymbol{\phi}^{0\prime} \delta^0\left(\mathbf{y}^\zeta\right), \tag{A.11}$$

$$\Delta \tilde{\mathbf{p}}^{\zeta,c}(\mathbf{y}) = \mathbf{\Upsilon}^{0\prime} \delta^0 \left(\mathbf{y}^\zeta \right), \tag{A.12}$$

where, using notation similar as in the proof of Proposition 1, $\delta^0 \left(\mathbf{y}^\zeta \right)$ is the $(|\mathscr{V}|-1) \times 1$ vector with lth element $p^0 \left(\mathbf{y}^\zeta \mid \mathbf{v}^l \right) - p^0 \left(\mathbf{y}^\zeta \mid \mathbf{v}^0 \right)$. It follows that

$$\bar{p} \left(\mathbf{y}^\zeta \mid \mathbf{v}^l \right) = p^0 \left(\mathbf{y}^\zeta \mid \mathbf{v}^0 \right) + \left[\boldsymbol{\phi}^0 + \mathbf{\Upsilon}^0 \bar{\mathbf{\Upsilon}}^{-1} \left[\mathbf{e}_l - \bar{\boldsymbol{\phi}} \right] \right]' \delta^0 \left(\mathbf{y}^\zeta \right). \tag{A.13}$$

Let $\bar{\mathbf{P}}_n^\zeta \left(\mathbf{v}^l \right)$ be the $(J+1)^n \times (J+1)$ matrix with (i,j)th element given by

$$\bar{p} \left(\mathbf{y}_n^{\zeta,i}, y_{n+1}^j \mid \mathbf{v}^l \right) = p^0 \left(\mathbf{y}_n^{\zeta,i}, y_{n+1}^j \mid \mathbf{v}^0 \right) + \left[\boldsymbol{\phi}^0 + \mathbf{\Upsilon}^0 \bar{\mathbf{\Upsilon}}^{-1} \left[\mathbf{e}_l - \bar{\boldsymbol{\phi}} \right] \right]' \delta^0 \left(\mathbf{y}_n^{\zeta,i}, y_{n+1}^j \right). \tag{A.14}$$

For $\bar{\boldsymbol{\theta}}$ to belong to the identified set Θ_{I}, it must satisfy rank $\left\{ \bar{\mathbf{P}}_n^\zeta \left(\mathbf{v}^l \right) \right\} = 1$, $\forall n \in \{1, \ldots, N-1\}$ and $\forall l \in \{0, \ldots, |\mathscr{V}|-1\}$. In particular, it must be the case that all minors of order 2 of $\bar{\mathbf{P}}_n^\zeta \left(\mathbf{v}^l \right)$ are equal to 0 $\forall n \in \{1, \ldots, N-1\}$ and $\forall l \in \{0, \ldots, |\mathscr{V}|-1\}$. Notice that each minor of order 2 defines a quadratic expression in $\boldsymbol{\phi}^0 + \mathbf{\Upsilon}^0 \bar{\mathbf{\Upsilon}}^{-1} \left[\mathbf{e}_l - \bar{\boldsymbol{\phi}} \right]$ with coefficients that depend on the true choice probabilities. Fix a specific $\mathbf{v}^l$ and consider the $N-1$ matrices $\bar{\mathbf{P}}_1^\zeta \left(\mathbf{v}^l \right), \ldots, \bar{\mathbf{P}}_{N-1}^\zeta \left(\mathbf{v}^l \right)$. From these $N-1$ matrices, one can construct $\sum_{n=1}^{N-1} \binom{n(J+1)}{2} \binom{J+1}{2}$ minors of order 2. For $\bar{\boldsymbol{\theta}}$ to belong to Θ_{I}, all the quadratic expressions corresponding to these minors must be equal to 0 when evaluated at $\bar{\boldsymbol{\theta}}$. A necessary (but not sufficient) condition for $\bar{\boldsymbol{\theta}} \in \Theta_{\mathrm{I}}$ is that the quadratic expressions must have a common solution, which defines restrictions on the true data generating process. To show that these restrictions are satisfied by a set of data generating processes that has Lebesgue measure zero within the space of choice probabilities, one can use the properties of multivariate resultants (see for instance Cox, Little, and O'Shea, 2005, Chap. 3). Since the quadratic expressions are not homogenous, one can homogenize them by multiplying some terms of the quadratic expressions with properly defined powers of an additional variable. This is without loss of generality since the resultant of the homogenized polynomials is the same as the resultant of their non-homogenized couterparts (e.g., Cox et al., 2005, p. 81). Since $\boldsymbol{\phi}^0 + \mathbf{\Upsilon}^0 \bar{\mathbf{\Upsilon}}^{-1} \left[\mathbf{e}_l - \bar{\boldsymbol{\phi}} \right]$ is a vector counting $|\mathscr{V}|-1$ elements, there are $|\mathscr{V}|$ unknowns in the homogenized polynomials. Pick $|\mathscr{V}|$ homogenous polynomials constructed from $|\mathscr{V}|$ minors of order 2, which is possible provided that $\sum_{n=1}^{N-1} \binom{n(J+1)}{2} \binom{J+1}{2} \geq |\mathscr{V}|$. By Cox et al. (2005), Thm. 2.3), these $|\mathscr{V}|$ homogenous polynomials in $|\mathscr{V}|$ unknowns have a common nontrivial solution if and only if the resultant of the $|\mathscr{V}|$ homogenous polynomials is equal to zero. The resultant is a uniquely defined polynomial in the coefficients of the homogenous polynomials and is therefore a polynomial in the choice probabilities $p^0 (\mathbf{y} \mid \mathbf{v})$ $\forall \mathbf{y} \in \mathscr{Y}^N$ and $\forall \mathbf{v} \in \mathscr{V}$ corresponding to the true data generating process. One can construct such a resultant for different subsets of $|\mathscr{V}|$ minors of order 2. The set of true choice probabilities for which all such resultants are simultaneously equal to 0 defines an algebraic variety (Allman et al., 2009, p. 3105). This variety is proper since any data generating process for which rank $\left\{ \bar{\mathbf{P}}_n^\zeta \left(\mathbf{v}^l \right) \right\} \neq 1$, for some $n \in \{1, \ldots, N-1\}$ and/or some $l \in \{0, \ldots, |\mathscr{V}|-1\}$ does not lie in it. It follows that the set of true choice probabilities for which all such resultants

are simultaneously equal to 0 is of dimension smaller than the size of the space of choice probabilities and hence the set of data generating processes for which $\bar{\boldsymbol{\theta}} \in \Theta_I$ necessary has Lebesgue measure zero within the space of choice probabilities $p(\mathbf{y}|\boldsymbol{\nu}) \; \forall \mathbf{y} \in \mathscr{Y}^N$ and $\forall \boldsymbol{\nu} \in \mathscr{V}$.

A.5. Proof of Proposition 4

Since both $\mathcal{Y}$ and $\mathcal{X}$ have finite and discrete support, $\mathbf{p}^0$ can be estimated as the vector of sample averages of properly defined dummy variables. Since $\hat{\mathbf{p}}$ is the only source of randomness in $g\left(\hat{\mathbf{p}}, \boldsymbol{\theta}\right)$, one can use the estimation and inference results from Chernozhukov et al. (2013) that are associated with parametrically estimated bounding functions. Chernozhukov et al. (2013), Thm. 4(a), p. 694) implies that

$$\limsup_{M \to \infty} \Pr\left(\inf_{l \in \mathscr{L}} g_l\left(\mathbf{p}^0, \boldsymbol{\theta}\right) \geq \hat{g}\left(\boldsymbol{\theta}, \alpha\right)\right) \leq \alpha. \tag{A.15}$$

When $\boldsymbol{\theta} \in \Theta_I$, $\inf_{l \in \mathscr{L}} g_l\left(\mathbf{p}^0, \boldsymbol{\theta}\right) \geq 0$, such that $\Pr\left(0 > \hat{g}\left(\boldsymbol{\theta}, \alpha\right)\right) \leq \Pr\left(\inf_{l \in \mathscr{L}} g_l\left(\mathbf{p}^0, \boldsymbol{\theta}\right) \geq \hat{g}\left(\boldsymbol{\theta}, \alpha\right)\right)$. By the definition of $\mathrm{CS}_M\left(\alpha\right)$ in (45), (A.15) implies that $\forall \boldsymbol{\theta} \in \Theta_I$:

$$\limsup_{M \to \infty} \Pr\left(\boldsymbol{\theta} \notin \mathrm{CS}_M\left(\alpha\right)\right) = \limsup_{M \to \infty} \Pr\left(0 > \hat{g}\left(\boldsymbol{\theta}, \alpha\right)\right) \leq \alpha. \tag{A.16}$$

Moreover, since $\mathrm{CS}_M\left(\alpha\right) = \emptyset \Rightarrow \boldsymbol{\theta} \notin \mathrm{CS}_M\left(\alpha\right)$ for any $\boldsymbol{\theta} \in \Theta_I$, it follows that $\Pr\left(\mathrm{CS}_M\left(\alpha\right) = \emptyset\right) \leq \Pr\left(\boldsymbol{\theta} \notin \mathrm{CS}_M\left(\alpha\right)\right)$. Therefore, for any $\boldsymbol{\theta} \in \Theta_I$:

$$\limsup_{M \to \infty} \mathrm{E}_{\boldsymbol{\theta}}\left[\xi_M\right] = \limsup_{M \to \infty} \Pr\left(\mathrm{CS}_M\left(\alpha\right) = \emptyset\right) \leq \limsup_{M \to \infty} \Pr\left(\boldsymbol{\theta} \notin \mathrm{CS}_M\left(\alpha\right)\right) \leq \alpha \tag{A.17}$$

which completes the proof.

APPENDIX B. IMPLEMENTATION OF THE INTERSECTION BOUNDS APPROACH

In the case with N players, $J+1$ actions and $|\mathscr{V}| = 2$ components in the mixture over the unobserved heterogeneity, the inequalities to be checked can be written as follows. The functions $Q\left(\mathbf{x}_E\right)$, $L_0\left(\mathbf{y}\right)$ and $L_1\left(\mathbf{y}\right)$ are defined in (30). The conditions in Proposition 2(i) can be rewritten as

$$Q\left(\mathbf{x}_E\right) + \frac{\phi}{\Upsilon} \geq 0 \text{ and } \frac{1-\phi}{\Upsilon} - Q\left(\mathbf{x}_E\right) \geq 0 \quad \forall \mathbf{x}_E \in \mathscr{X}_E. \tag{B.1}$$

Similarly, the conditions in Proposition 2(ii) can be rewritten as

$$-\frac{\phi}{\Upsilon} - \min\left\{L_0\left(\mathbf{y}\right), L_1\left(\mathbf{y}\right)\right\} \geq 0 \text{ and } \max\left\{L_0\left(\mathbf{y}\right), L_1\left(\mathbf{y}\right)\right\} - \frac{1-\phi}{\Upsilon} \geq 0 \quad \forall \mathbf{y} \in \mathscr{Y}^N. \tag{B.2}$$

Finally, the conditions in Proposition 2(iii) can be written in terms of all the possible minors of order 2 computed from the matrices $\mathbf{P}_n^{\zeta}\left(\boldsymbol{\nu}^0\right)$ and $\mathbf{P}_n^{\zeta}\left(\boldsymbol{\nu}^1\right)$ for $n = 1, \ldots, N-1$. There are $K(n) = \binom{n(J+1)}{2}\binom{J+1}{2}$ minors of order 2 for each such matrix. When $|\mathscr{V}| = 2$, the (i,j)th

element of $\mathbf{P}_n^\zeta\left(\boldsymbol{v}^0\right)$ and $\mathbf{P}_n^\zeta\left(\boldsymbol{v}^1\right)$ are, respectively,

$$p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j\,\middle|\,\boldsymbol{v}^0\right)=p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j\,\middle|\,\mathbf{x}_E^0\right)-\frac{\phi}{\Upsilon}\left[p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j\,\middle|\,\tilde{\mathbf{x}}_E^1\right)-p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j\,\middle|\,\mathbf{x}_E^0\right)\right],$$
$$\text{(B.3)}$$

$$p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j\,\middle|\,\boldsymbol{v}^1\right)=p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j\,\middle|\,\mathbf{x}_E^0\right)+\frac{1-\phi}{\Upsilon}\left[p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j\,\middle|\,\tilde{\mathbf{x}}_E^1\right)-p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j\,\middle|\,\mathbf{x}_E^0\right)\right].$$
$$\text{(B.4)}$$

Each minor of order 2 involves exactly two different values of $\mathbf{y}_n^\zeta$ and two different values of y_{n+1}. Let $\mathbf{P}_n^\zeta\left(\boldsymbol{v}^0\right)$'s kth minor of order 2 be

$$p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j\,\middle|\,\boldsymbol{v}^0\right)p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^t\,\middle|\,\boldsymbol{v}^0\right)-p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^j\,\middle|\,\boldsymbol{v}^0\right)p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^t\,\middle|\,\boldsymbol{v}^0\right). \tag{B.5}$$

Using (B.3), one can write the condition that this kth minor of order 2 is equal to 0, using the following inequalities:

$$M_0\left(k,n\right)\equiv a_2\left(k,n\right)\left[\frac{-\phi}{\Upsilon}\right]^2+a_1\left(k,n\right)\left[\frac{-\phi}{\Upsilon}\right]+a_0\left(k,n\right)\geq0\quad\text{and}\quad-M_0\left(k,n\right)\geq0,$$
$$\text{(B.6)}$$

where,

$$a_0\left(k,n\right)\equiv p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j|\mathbf{x}_E^0\right)p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^t|\mathbf{x}_E^0\right)-p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^j|\mathbf{x}_E^0\right)p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^t|\mathbf{x}_E^0\right),$$
$$\text{(B.7)}$$

$$\begin{aligned}a_1\left(k,n\right)\equiv&\,p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j|\mathbf{x}_E^0\right)\left[p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^t|\tilde{\mathbf{x}}_E^1\right)-p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^t|\mathbf{x}_E^0\right)\right]\\&+p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^t|\mathbf{x}_E^0\right)\left[p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j|\tilde{\mathbf{x}}_E^1\right)-p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j|\mathbf{x}_E^0\right)\right]\\&-p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^j|\mathbf{x}_E^0\right)\left[p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^t|\tilde{\mathbf{x}}_E^1\right)-p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^t|\mathbf{x}_E^0\right)\right]\\&-p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^t|\mathbf{x}_E^0\right)\left[p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^j|\tilde{\mathbf{x}}_E^1\right)-p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^j|\mathbf{x}_E^0\right)\right],\end{aligned}$$
$$\text{(B.8)}$$

$$\begin{aligned}a_2\left(k,n\right)\equiv&\left[p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j|\tilde{\mathbf{x}}_E^1\right)-p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^j|\mathbf{x}_E^0\right)\right]\left[p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^t|\tilde{\mathbf{x}}_E^1\right)-p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^t|\mathbf{x}_E^0\right)\right]\\&-\left[p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^t|\tilde{\mathbf{x}}_E^1\right)-p\left(\mathbf{y}_n^{\zeta,i},y_{n+1}^t|\mathbf{x}_E^0\right)\right]\left[p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^j|\tilde{\mathbf{x}}_E^1\right)-p\left(\mathbf{y}_n^{\zeta,s},y_{n+1}^j|\mathbf{x}_E^0\right)\right].\end{aligned}$$
$$\text{(B.9)}$$

Similarly, the inequality conditions for $\mathbf{P}_n^\zeta\left(\boldsymbol{v}^1\right)$'s kth minor of order 2 being equal to 0 are

$$M_1\left(k,n\right)\equiv a_2\left(k,n\right)\left[\frac{1-\phi}{\Upsilon}\right]^2+a_1\left(k,n\right)\left[\frac{1-\phi}{\Upsilon}\right]+a_0\left(k,n\right)\geq0\quad\text{and}\quad-M_1\left(k,n\right)\geq0.$$
$$\text{(B.10)}$$

In order to ensure that the condition in Proposition 4(ii) is satisfied, one cannot directly include all the restrictions written above. In particular, notice that $Q\left(\mathbf{x}_E^0\right)=0$ and $Q\left(\tilde{\mathbf{x}}_E^1\right)=$

1, which implies that conditions (B.1) for these values of $\mathbf{x}_E$ are equivalent to

$$\phi \geq 0; \quad 1-\phi \geq 0; \quad \phi + \Upsilon \geq 0; \text{ and } 1-\phi-\Upsilon \geq 0. \tag{B.11}$$

Notice that the conditions in (B.11) do not depend on $\mathbf{p}$ and would therefore violate Proposition 4(ii). Nonetheless, (B.11) restricts the parameter space Θ over which the test is performed. In the implementation of the test, the problematic conditions from $\mathbf{g}(\mathbf{p}, \boldsymbol{\theta})$ are dropped, but Θ is restricted to $\boldsymbol{\theta}$'s that satisfy (B.11).

To sum up, the vector $\mathbf{g}(\mathbf{p}, \boldsymbol{\theta})$ includes the following $L=6$ inequalities:

$$\min_{\mathbf{x}_E \in \mathscr{X}_E \setminus \left\{\mathbf{x}_E^0, \tilde{\mathbf{x}}_E^1\right\}} Q(\mathbf{x}_E) + \frac{\phi}{\Upsilon} \geq 0, \tag{B.12}$$

$$\frac{1-\phi}{\Upsilon} - \max_{\mathbf{x}_E \in \mathscr{X}_E \setminus \left\{\mathbf{x}_E^0, \tilde{\mathbf{x}}_E^1\right\}} Q(\mathbf{x}_E) \geq 0, \tag{B.13}$$

$$-\frac{\phi}{\Upsilon} - \max_{\mathbf{y} \in \mathscr{Y}^N} \{\min\{L_0(\mathbf{y}), L_1(\mathbf{y})\}\} \geq 0, \tag{B.14}$$

$$\min_{\mathbf{y} \in \mathscr{Y}^N} \{\max\{L_0(\mathbf{y}), L_1(\mathbf{y})\}\} - \frac{1-\phi}{\Upsilon} \geq 0, \tag{B.15}$$

$$\min_{\substack{n=1,\ldots,N-1 \\ k=1,\ldots,K(n)}} \{\min\{M_0(k,n), M_1(k,n)\}\} \geq 0, \tag{B.16}$$

$$\min_{\substack{n=1,\ldots,N-1 \\ k=1,\ldots,K(n)}} \{\min\{-M_0(k,n), -M_1(k,n)\}\} \geq 0, \tag{B.17}$$

and only values of $\boldsymbol{\theta}$ satisfying (B.11) are considered when constructing the confidence set.

The test is implemented by discretizing Θ into a fine grid and checking whether each $\boldsymbol{\theta}$ in this discretized space belong to $\text{CS}_M(\alpha)$. The null hypothesis is rejected if there is no such $\boldsymbol{\theta}$ that belongs to $\text{CS}_M(\alpha)$. In order to check if $\boldsymbol{\theta} \in \text{CS}_M(\alpha)$, Algorithm 1 described in Chernozhukov et al. (2013), p. 708) is used. It is reproduced here for completeness.

1. Set $\eta_M \equiv 1 - 0.1/\ln[M]$ and simulate R independent draws $\mathbf{z}_1, \ldots, \mathbf{z}_R$ from the $\dim(\mathbf{p})$-variate standard normal distribution.
2. Compute $\hat{\boldsymbol{\Sigma}}$.
3. For each $l \in \mathscr{L}$, compute $\hat{\mathbf{j}}_l(\boldsymbol{\theta})' \equiv \dfrac{\partial g_l(\hat{\mathbf{p}}, \boldsymbol{\theta})}{\partial \mathbf{p}'} \hat{\boldsymbol{\Sigma}}^{1/2}$ and set $\hat{\sigma}_l(\boldsymbol{\theta}) \equiv \left\|\hat{\mathbf{j}}_l(\boldsymbol{\theta})\right\|/\sqrt{M}$, where $\|\cdot\|$ is the euclidean norm.
4. Compute

$$\hat{k}(\boldsymbol{\theta}, \eta_M) = \eta_M\text{-quantile of } \left\{\sup_{l \in \mathscr{L}} \left(\hat{\mathbf{j}}_l(\boldsymbol{\theta})' \mathbf{z}_r / \left\|\hat{\mathbf{j}}_l(\boldsymbol{\theta})\right\|\right), r = 1, \ldots, R\right\} \tag{B.18}$$

and

$$\hat{\mathscr{L}}(\boldsymbol{\theta}) = \left\{l \in \mathscr{L} : \hat{g}_l(\hat{\mathbf{p}}, \boldsymbol{\theta}) \leq \min_{l \in \mathscr{L}} \left(\hat{g}_l(\hat{\mathbf{p}}, \boldsymbol{\theta}) + \hat{k}(\boldsymbol{\theta}, \eta_M) \hat{\sigma}_l(\boldsymbol{\theta})\right) + 2\hat{k}(\boldsymbol{\theta}, \eta_M) \hat{\sigma}_l(\boldsymbol{\theta})\right\}. \tag{B.19}$$

5. Compute

$$\hat{c}(\boldsymbol{\theta}, 1-\alpha) = (1-\alpha)\text{-quantile of } \left\{ \sup_{l \in \hat{\mathcal{L}}(\boldsymbol{\theta})} \left(\hat{\mathbf{j}}_l(\boldsymbol{\theta})' \mathbf{z}_r / \left\| \hat{\mathbf{j}}_l(\boldsymbol{\theta}) \right\| \right), r = 1, \ldots, R \right\}$$

(B.20)

and set $\hat{g}(\boldsymbol{\theta}, \alpha) = \inf_{l \in \mathcal{L}} \left\{ g_l(\hat{\mathbf{p}}, \boldsymbol{\theta}) + \hat{c}(\boldsymbol{\theta}, 1-\alpha) \hat{\sigma}_l(\boldsymbol{\theta}) \right\}$.

One rejects $\boldsymbol{\theta} \in \Theta_I$ if $\hat{g}(\boldsymbol{\theta}, \alpha) < 0$.

To reduce the computational burden of the grid search, one should leverage the fact that the conditions in Proposition 2(iii) drastically shrink the identified set. The simulations therefore use as starting values of $\boldsymbol{\theta}$ the values on the grid which are closest to the ones that solve (B.6) and (B.10).

REFERENCES

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

Aguirregabiria, V. & P. Mira (2019) Identification of games of incomplete information with multiple equilibria and unobserved heterogeneity. *Quantitative Economics* 10(4), 1659–1701.

Allman, E., C. Mathias, & J. Rhodes (2009) Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics* 37(6A), 3099–3132.

Andrews, D. & 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. & 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. (2020) The econometrics of static games. *Annual Review of Economics* 12(3), 1–31.

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

Athey, S. & P. Haile (2002) Identification of standard auction models. *Econometrica* 70(6), 2107–2140.

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

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

Bajari, P., H. Hong, & D. Nekipelov (2013) Game theory and econometrics: A survey of some recent research. In D. Acemoglu, M. Arellano, & E. Dekel (eds.), *Advances in Economics and Econometrics, 10th World Congress*, pp. 3–52. Cambridge.

Ben-Porath, E. (1992) Signaling future actions and the potential for sacrifice. *Journal of Economic Theory* 57(1), 36–51.

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

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

Berry, S. & E. Tamer (2006) Identification in models of oligopoly entry. *Advances in Economics and Econometrics* 2, 46–85.

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

Brock, W. & S. Durlauf (2001) Discrete choice with social interactions. *Review of Economic Studies* 68(2), 235–260.

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

Campo, S., I. Perrigne, & Q. Vuong (2003) Asymmetry in first-price auctions with affiliated private values. *Journal of Applied Econometrics* 18(2), 179–207.

Chan, K. (2006) A simple mathematical approach for determining intersection of quadratic surfaces. *Multiscale Optimization Methods and Applications*, 271–298.

Chernozhukov, V., S. Lee, & A. Rosen (2013) Intersection bounds: Estimation and inference. *Econometrica* 81(2), 667–737.

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

Cox, D., J. Little, & D. O'Shea (2005) Using Algebraic Geometry, Graduate Texts in Mathematics. Springer.

Dawid, A. (1979) Conditional independence in statistical theory. *Journal of the Royal Statistical Society - Series B* 41(1), 1–31.

De Paula, A. (2013) Econometric analysis of games with multiple equilibria. *Annual Review of Economics* 5(1), 107–131.

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

de Paula, A. & X. Tang (2020) Testable implications of multiple equilibria in discrete games with correlated types. Cemmap Working paper, pp. 1–35. https://www.cemmap.ac.uk/publication/testable-implications-of-multiple-equilibria-in-discrete-games-with-correlated-types/.

Delgado, M. & J. Hidalgo (2000) Nonparametric inference on structural breaks. *Journal of Econometrics* 96(1), 113–144.

Espin-Sanchez, J., A. Parra, & Y. Wang (2022) Equilibrium uniqueness in entry games with private information. *RAND Journal of Economics* 50(3), 568–590.

Galichon, A. & 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.

Gowrisankaran, G. & J. Krainer (2011) Entry and pricing in a differentiated products industry: Evidence from the ATM market. *RAND Journal of Economics* 42(1), 1–22.

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

Guerre, E., I. Perrigne, & Q. Vuong (2009) Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. *Econometrica* 77(4), 1193–1227.

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

Haile, P., H. Hong, & M. Shum (2003) Nonparametric tests for common values at first-price sealed-bid auctions. NBER Working paper.

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

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

Hong, H. & M. Shum (2002) Increasing competition and the winner's curse: Evidence from procurement. *Review of Economic Studies* 69(4), 871–898.

Hsu, Y., C. Liu, & X. Shi (2019) Testing generalized regression monotonicity. *Econometric Theory* 35(6), 1146–1200.

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.

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

Igami, M. & N. Yang (2016) Unobserved heterogeneity in dynamic games: Cannibalization and preemptive entry of hamburger chains in Canada. *Quantitative Economics* 7(2), 483–521.

Jia, P. (2008) What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica* 76(6), 1263–1316.

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

Kasahara, H. & K. Shimotsu (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. & 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.

Kline, B. & E. Tamer (2016) Bayesian inference in a class of partially identified models. *Quantitative Economics* 7(2), 329–366.

Lewbel, A. & X. Tang (2015) Identification of games with incomplete information using excluded regressors. *Journal of Econometrics* 189(1), 229–244.

Luo, Y., P. Xiao, & R. Xiao (2022) Identification of dynamic games with unobserved heterogeneity and multiple equilibria. *Journal of Econometrics* 226(2), 343–367.

Magesan, A. (2018) Identification of static and dynamic games of incomplete information with multiple equilibria in the data. SSRN Working paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2985097.

Magnolfi, L. & C. Roncoroni (2022) Estimation of discrete games with weak assumptions on information. *The Review of Economic Studies*. https://academic.oup.com/restud/advance-article/doi/10.1093/restud/rdac058/6670639?login=true.

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

Mourifié, I., M. Henry, & R. Méango (2020) Sharp bounds and testability of a Roy model of STEM major choices. *Journal of Political Economy* 128(8).

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

Muller, H. (1992) Change-points in nonparametric regression analysis. *Annals of Statistics* 20(2), 737–761.

Navarro, S. & Y. Takahashi (2012) A semiparametric test of agent's information sets for games of incomplete information. EconStor Working paper. https://www.econstor.eu/handle/10419/94083?locale=en.

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

Pakes, A., M. Ostrovsky, & 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. & P. Schmidt-Dengler (2008) Asymptotic least squares estimators for dynamic games. *Review of Economic Studies* 75(3), 901–928.

Seim, K. (2006) An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics* 37(3), 619–640.

Shi, X. & 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.

Syrgkanis, V., E. Tamer, & J. Ziani (2018) Inference on auctions with weak assumptions on information. Arxiv Working paper. https://arxiv.org/abs/1710.03830.

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

Van Damme, E. (1989) Stable equilibria and forward induction. *Journal of Economic Theory* 48(2), 476–496.

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