Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 11-003

"Inference of Signs of Interaction Effects In Simultaneous Games with Incomplete Information" Second Version

by

Aureo de Paula and Xun Tang

http://ssrn.com/abstract=1761377

Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information*

Áureo de Paula †

University of Pennsylvania

Xun Tang ‡

University of Pennsylvania

This version: February 12, 2011

*We thank Andres Aradillas-Lopez, Federico Bugni, Ivan Canay, Andrew Chesher, Steve Durlauf, George Deltas, Yanqin Fan, Hanming Fang, Paul Grieco, Phil Haile, Jim Heckman, Bo Honoré, Jean-Francois Houde, Steffen Huck, John Kennan, Ivana Komunjer, Tong Li, Chuck Manski, Salvador Navarro, Aviv Nevo, Rob Porter, Seth Richards, Adam Rosen, Andres Santos, Frank Schorfheide, Azeem Shaikh, Matt Shum, Kevin Song, Yixiao Sun, Elie Tamer, Yuanyuan Wan, Haiqing Xu and participants at CIREQ's Conference on the Econometrics of Social Interactions (October, 2009), Econometric Society World Congress 2010, SEA Meetings 2010, National University of Singapore, Northwestern, Princeton, University of Chicago, UCSD, University of Illinois (Urbana-Champaign), Vanderbilt, Wisconsin (Madison) and UCL for helpful comments. We owe special thanks to Andrew Sweeting for kindly providing us with the data and for helpful suggestions.

†Dept of Economics, U Penn, Philadelphia, PA 19104. E-mail: aureo@sas.upenn.edu

‡Dept of Economics, U Penn, Philadelphia, PA 19104. E-mail: xuntang@sas.upenn.edu

Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information

Abstract

This paper studies the inference of interaction effects (impacts of players' actions on each other's payoffs) in discrete simultaneous games with incomplete information. We propose an easily implementable test for the signs of state-dependent interaction effects that does not require parametric specifications of players' payoffs, the distributions of their private signals or the equilibrium selection mechanism. The test relies on the commonly invoked assumption that players' private signals are independent conditional on observed states. The procedure is valid in (but does not rely on) the presence of multiple equilibria in the data-generating process (DGP). As a by-product, we propose a formal test for multiple equilibria in the DGP. We also show how to extend our arguments to identify signs of interaction effects when private signals are correlated. We provide Monte Carlo evidence of the test's good performance in finite samples. We then implement the test using data on radio programming of commercial breaks in the U.S., and infer stations' incentives to synchronize their commercial breaks. Our results support the earlier finding by Sweeting (2009) that stations have stronger incentives to coordinate and air commercials at the same time during rush hours and in smaller markets.
JEL Codes: C01, C72

1 Introduction

Strategic interaction effects occur when a player's action choice affects not only his or her own payoff but also those of other players. In simultaneous discrete games of incomplete information, each person has a private signal about his or her payoff, while the joint distribution of such private signals is common knowledge among all players.[1] In a Bayesian Nash equilibrium (BNE), individuals act to maximize their expected payoffs given their knowledge of these distributions and the payoff structure. Such models have found applications in a variety of empirical contexts where players are uncertain about their competitors' payoffs given their own information. These include, for example, airing commercials at radio stations (Sweeting (2009)) and peer effects in recommendations by financial analysts (Bajari, Hong, Krainer, and Nekipelov (2010)).

Earlier works have studied the identification and estimation of these games using a wide spectrum of restrictions. These include (but are not limited to) the independence of private signals from observable covariates, parametric specification of relevant distributions or utility functions, or constraints on the set of Bayesian Nash equilibria. In comparison, we focus on inference of the signs of interaction effects, which are allowed to be individual-specific and state-dependent, under a minimal set of nonparametric restrictions on private signals and payoff structures. Our choice of focus is motivated by two considerations. First, signs of interaction effects alone may have important policy implications. For example, if agents have an incentive to coordinate on a particular action then an exogenous intervention that induces a subset of participants to choose a certain action should at the same time also incentivize

[1]Recent work by Grieco (2010) studies a class of games with flexible information structures that also subsume games with complete information where players know each other's payoffs for sure. In a similar spirit, Navarro and Takahashi (2009) suggest a test for the information structure that, among other things, relies on a degenerate equilibrium selection rule and independence between residuals and observed covariates. Other papers have also dealt with unobserved heterogeneity across games which is observed by players but not econometricians (e.g. Sweeting (2009), Aguirregabiria and Mira (2007) and Arcidiacono and Miller (2010), these last two in a dynamic setting).

other players to act accordingly. Second, while point identification and estimation of the full structure of such games inevitably hinge on parametric restrictions, inference on signs of interaction effects can be done under minimal nonparametric restrictions on the structure. Such inference is valid even in the presence of multiple equilibria and does not invoke any assumptions on the equilibrium selection mechanism in the data-generating process. This is particularly notable, since almost all previous work has relied on stringent assumptions about equilibrium selection or multiplicity to attain identification (e.g., the single-equilibrium assumption in Bajari, Hong, Krainer, and Nekipelov (2010) and Tang (2010), equilibrium uniqueness in Seim (2006) or Aradillas-Lopez (2010), the restriction to monotone, threshold-crossing Bayesian-Nash equilibria of Wan and Xu (2010) or the symmetry of equilibria and payoff functions and parametrization of equilibrium selection mechanism as in Sweeting (2009)).[2]

The existence of multiple equilibria in the data can be exploited to infer *the signs* of strategic interactions. If players' private signals are independent from each other given observed covariates, then their chosen actions must be uncorrelated in any single equilibrium. On the other hand, if multiple equilibria exist in the data, then the joint distribution of actions observed is a mixture of those implied in each single equilibrium. This leads to correlations between the players' actions observed from data. We show in Section 3 that signs of correlations between players' actions are determined by the signs of the strategic interaction effects. As a by-product, the correlations also allow us to identify the existence of multiple equilibria in the data (see below). The assumption of conditional independence of private information is commonly maintained in the literature on estimation and inference in static games with incomplete information (see, for example, Seim (2006), Aradillas-Lopez

[2]As indicated in Berry and Tamer (2007), another possibility is to resort to partial identification. Examples of such a strategy in games of *complete information* are Beresteanu, Molchanov, and Molinari (2009), Ciliberto and Tamer (2009), Galichon and Henry (2009) and earlier references cited in Berry and Tamer (2007). Also in games of *complete information*, Bjorn and Vuong (1984) parameterize the equilibrium selection mechanism.

(2010), Berry and Tamer (2007), Bajari, Hong, Krainer, and Nekipelov (2010), Bajari, Hahn, Hong, and Ridder (forthcoming), Brock and Durlauf (2007), Sweeting (2009) and Tang (2010)).[3] The assumption can also be found in the literature on the estimation of dynamic games with incomplete information.

We also generalize these arguments for identifying the signs of interaction effects to allow for the possibility that in the data there is only a unique equilibrium for a given state. The idea relies on the following simple intuition. Suppose that for some player i there exists a sub-vector of state variables that affect other players' payoffs or private signals but not his or her own. Then sign of the correlation between actions chosen by i and others across different realizations of such "excluded" states must be solely determined by others' actions affect i's payoffs, provided the private signals are independent given observed states. Such exclusion restrictions on state variables arise naturally in many applications and have been used before in similar contexts.

Another contribution of this paper is to introduce a formal test for the presence of multiple equilibria in the data-generating process. Testing for multiple equilibria is of practical importance in empirical research, because existing estimation methods often rely on the occurrence of a single equilibrium in the data. The test we propose is a natural outcome of the logic used in our inference of the signs of interaction effects. An innovation of our test for multiple equilibria is to use a stepwise multiple testing procedure to infer whether *each individual player* has different strategies across the multiple equilibria in the data-generating process. This is particularly interesting for structural estimation of games involving three or more players, in which a subset of players may stick to the same strategy across multiple equilibria. Semiparametric methods based on the assumption of a unique equilibrium can still be applied to consistently estimate payoff parameters for those players

[3]In a subsection, Aradillas-Lopez (2010) suggests an estimation procedure to handle cases in which the assumption is violated, but relies on the assumption that a single equilibrium is played in the data. Another exception is Wan and Xu (2010) who nevertheless also require that a unique (monotone) Bayesian-Nash equilibrium be played in the data.

who do not switch between strategies in multiple equilibria. Hence, it is useful to infer the *identity* of such players from observed distributions of actions. Our test is known to effectively control the probability of rejecting at least one of the true single null hypotheses.

We also provide identification results that relax the conditional independence of private signals. Such an extension relies on econometricians' observation of groups (or clusters) of games within which players follow strategies prescribed by the same BNE. For example, a market or household observed over multiple periods or a cluster of games from similar cultural traits or geographic region could often be justified as one such group. Within a given group, permuting players across independent games still leads to observations where the same equilibrium is played. Provided games are independent within groups, private signals of players across these games will be independent, and permuted versions of the games would mimic the conditional independence assumption.

Finally, we apply our methodology to investigate radio stations' incentives to coordinate on commercial breaks using the data from Sweeting (2009). Relaxing the parametric and symmetry assumptions in that paper, we confirm his findings that incentives to coordinate are stronger during rush hour and in smaller markets.

The paper proceeds as follows. We present our basic model and empirical characterization in the next section. In Section 3, we present the main results on the identification of the sign of interaction effects. Section 4 outlines general testing procedures for inference. We generalize our results in section 5. Monte Carlo experiments and an application to joint retirement are presented in Sections 6 and 7. Section 8 concludes.

2 The Model and Empirical Context

We consider a simultaneous discrete game with incomplete information involving N players. Each player i chooses an action D_i from two alternatives, $\{1, 0\}$. A vector of states $X \in \mathbb{R}^K$ is common knowledge among all players. A vector of private information (or "types/signals")

$\epsilon \equiv (\epsilon_i)_{i \leq N} \in \mathbb{R}^N$ is such that ϵ_i is only observed by player i. Throughout the paper we will use upper case letters for random variables and lower case for their realized values. We use Ω_R to denote the support of any generic random vector $R = (R_1, R_2)$, and let F_R, $F_{R_1|R_2}$ denote respectively the marginal and conditional distributions in the data-generating process (DGP). Conditional on a given state $X = x$, private information ϵ is jointly distributed according to the CDF $F_{\epsilon|X}(\cdot|x)$. The payoff for player i from choosing action 1 is $U_{1i}(X, \epsilon_i) \equiv u_i(X) + (\sum_{j \neq i} D_j)\delta_i(X) - \epsilon_i$, while the return from the other action $U_{0i}(X, \epsilon_i)$ is normalized to 0. For example, consider a game with two players and a payoff structure such that player i obtains $U_{ai}(x) \equiv \tilde{u}_{ai}(x) + \tilde{\delta}_{ai}(x)\mathbf{1}(j \text{ plays } a) - \tilde{\epsilon}_{ai}$ if she plays $a \in \{0, 1\}$. In this case, our analysis focusses on $u_i(x) \equiv \tilde{u}_{1i}(x) - \tilde{u}_{0i}(x)$, $\delta_i(x) \equiv \tilde{\delta}_{1i}(x) - \tilde{\delta}_{0i}(x)$ and $\epsilon_i \equiv \tilde{\epsilon}_{1i} - \tilde{\epsilon}_{0i}$, since decisions will depend only on the differences of payoffs. Intuitively, $u_i(X)$ specifies a base return from action 1 for player i. Meanwhile $\delta_i(X)$ captures interaction effects on i's payoff due to another player j who chooses 1. The return functions $(u_i, \delta_i)_{i=1}^N$ and the distribution of private information $F_{\epsilon|X}$ are common knowledge among all players. We maintain the following identifying restrictions on $F_{\epsilon|X}$ throughout the paper.

Assumption 1 *Conditional on any $x \in \Omega_X$, ϵ_i is independent of $(\epsilon_j)_{j \neq i}$ for all i and has positive density over $\mathbb{R}^N$.*

Assumption 1 allows X to be correlated with private information of the players, as is plausible in empirical applications. This conditional independence restriction is commonly used in the estimation literature for both static and dynamic games with incomplete information. A pure strategy for player i in this Bayesian game is a mapping $s_i : \Omega_{X,\epsilon_i} \rightarrow \{0, 1\}$. Letting $S_i(X, \epsilon_i)$ denote an equilibrium strategy for player i, the equilibrium behavior prescribes:

$$S_i(X, \epsilon_i) = \begin{cases} 1, & \text{if } u_i(X) + \delta_i(X)\Sigma_{j \neq i}\mathbb{E}\left[S_j(X, \epsilon_j)|X, \epsilon_i\right] - \epsilon_i \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Under Assumption 1, $\mathbb{E}\left[S_j(X, \epsilon_j)|X = x, \epsilon_i\right] = \mathbb{E}\left[S_j(X, \epsilon_j)|X = x\right] \equiv p_j(x)$, and a Bayesian Nash equilibrium (BNE) in pure strategies (given state x) can be characterized by a profile of choice probabilities $p(x) \equiv [p_1(x), \ldots, p_N(x)]$ such that for all $x \in \Omega_X$,

$$p_i(x) = F_{\epsilon_i|X=x}\left(u_i(x) + \delta_i(x)\Sigma_{j \neq i} p_j(x)\right) \text{ for all } i = 1, \ldots, N \tag{1}$$

where $p_i(x)$ is player i's probability of choosing action 1 conditional on the state x and $F_{\epsilon_i|X}$ is the marginal distribution of ϵ_i conditional on X. Let $\mathcal{L}_{x,\theta}$ denote the set of BNE (as summarized by solutions in p in (1)) for a given x and structure $\theta \equiv \{(u_i, \delta_i)_{i=1,\ldots,N}, F_{\epsilon|X}\}$. The existence of pure-strategy BNE for any given x follows from Brouwer's Fixed Point Theorem and the continuity of $F_{\epsilon_i|X}$ under Assumption 1. In general there may be multiple BNE, depending on the specifications of $F_{\epsilon|X}$, u_i and δ_i.

The model specification rules out general heterogeneous interaction effects that may vary with the identities of each pair of competing players (e.g. δ_{ij}). Nonetheless, we can extend our inference approach to allow players' payoffs to be affected by competitors' decisions in general forms that are known to researchers (see discussions in Section 3). This would be the case, for example, if payoffs depends on the proportion (instead of the sum) of agents taking an action, or on the action of at least one other person (but not on the action of additional agents beyond that) (i.e. $f_i(x, D_{-i}) = \max_{j \neq i}(D_j)$), or even if it change only when all competitors take a particular action (i.e. $f_i(x, D_{-i}) = \min_{j \neq i}(D_j)$).

This model differs qualitatively from the social interaction model studied in Brock and Durlauf (2007) and that in Sweeting (2009) in that it allows for asymmetry in players' payoff functions and equilibria. Thus, even when payoffs are symmetric, we allow for asymmetric BNE where the implied choice probabilities could vary across players, and multiple asymmetric BNE can arise regardless of the signs of interaction effects. This makes the task of detecting multiple BNE and signs of interaction effects more interesting as well as more challenging.

We assume econometricians have access to a large cross-section of independent games between N players. In each game, they observe choices of actions by all players and realized states x, but do not observe $(\epsilon_i)_{i \leq N}$ or know the form of $(u_i, \delta_i)_{i \leq N}$ and $F_{\epsilon|X}$. Our analysis posits (i) that the structure ($(u_i, \delta_i)_{i \leq N}$ and $F_{\epsilon|X}$) is fixed across all games observed, and (ii) that the choice data observed is generated by players following the pure strategies prescribed by BNE. Econometricians are interested in learning (at least some features of) the structure $(u_i, \delta_i)_{i \leq N}$ and $F_{\epsilon|X}$ from the observable joint distribution of X and $(D_i)_{i \leq N}$.

Suppose the choices observed in the data are known to be generated from a single BNE in the DGP for all $x \in \Omega_X$. This may arise because either (a) the solution to (1) is unique, or (b) the system of equations in (1) admits multiple solutions but the equilibrium selection in the DGP is degenerate in one of the multiple solutions. Then (1) offers a link between observable conditional choice patterns and structural elements $(u_i, \delta_i)_{i \leq N}, F_{\epsilon|X}$. Estimation can then be done under various restrictions on u, δ and $F_{\epsilon|X}$ (see Aradillas-Lopez (2010), Berry and Tamer (2007), Bajari, Hong, Krainer, and Nekipelov (2010) and Tang (2010) for more details). We say there are *multiple BNE in DGP* if there are several solutions to (1) and the equilibrium selection mechanism in the data is *not* degenerate at any one of them.

This link between observed choice patterns and structural elements may nonetheless break down when there are multiple equilibria in the data-generating process. To see this, let $\Lambda_{x,\theta}$ be an equilibrium selection mechanism (i.e. a distribution over $\mathcal{L}_{x,\theta}$) in the data-generating process that may depend on x and θ, but is independent from the vector of private information $(\epsilon_i)_{i \leq N}$. That Λ depends on x but not realizations of ϵ_i captures the idea that only information commonly known to all players may plausibly affect which equilibrium is played in the data-generating process (see Myerson (1991), pp.371-2).

To simplify the notation, we drop subscripts θ from $\Lambda_{x,\theta}$ and $\mathcal{L}_{x,\theta}$ in the subsequent sections when there is no ambiguity. For any x such that $\mathcal{L}_x$ is not a singleton, the conditional choice probability observed in the data is a mixture of the conditional choice probabilities implied by each pure-strategy BNE in $\mathcal{L}_x$. That is, $p_i^*(x) = \int_{\mathcal{L}_x} p_i^l d\Lambda_x(p^l)$, where $p_i^*(x)$ is the actual marginal probability that i chooses 1 conditional on x observed from data, and $p^l \equiv (p_i^l)_{i \leq N}$ is a generic element in the set of possible BNE $\mathcal{L}_x$, with l indexing the equilibria in $\mathcal{L}_x$ and p_i^l, the marginal probability for i to choose 1 given x (and the structure θ) implied in equilibrium l. While, by definition, the fixed point characterization in (1) holds for every single BNE $p^l \in \mathcal{L}_x$, it does not necessarily hold for the vector of mixture marginals $p^* \equiv (p_i^*)_{i \leq N}$ observed.

Researchers have taken different approaches to deal with the issue of multiple equilibria in empirical work. Each of these strategies (which can also be combined) has some limitations. We are interested in constructing a robust way to test for the existence of multiple equilibria and to recover the sign of interactions under minimum restrictions on the model primitives.

3 Identifying Signs of Interaction Effects

3.1 The basic idea

We now show how to detect the presence of multiple BNE in the data-generating process and identify signs of interaction effects $\delta_i(x)$ for any i given any x. The sign reveals the nature of strategic incentives among players. Compared with earlier works, our sign identification has several innovations and contributions. First, our test does not invoke any parametric restrictions on players' preferences or distributions of private information. Second, it allows the strategic incentives (as captured by the sign of δ_i) to be a function of states x. Third, our approach is robust to the presence of multiple BNE. In fact, while the existence of multiple BNE at first precludes complete identification of the structure, it does help identify the sign of interaction effects. This possibility is informally outlined, for example, in Manski (1993) and in Sweeting (2009).[4]

We first show how to detect multiple BNE in the data using observed distributions. Define

$$\gamma_i^l(x) \equiv \mathbb{E}_l \left(\Sigma_{j \neq i} D_j | X = x \right) = \Sigma_{j \neq i} p_j^l(x)$$

where $\mathbb{E}_l$ denotes the expectation with respect to the distribution of $(D_i)_{i \leq N}$ induced in the

[4] "The prospects for identification may improve if $f(\cdot, \cdot)$ is non-linear in a manner that generates multiple social equilibria" (p. 539, Manski (1993)).

equilibrium $p^l \in \mathcal{L}_x$. Define $sign(a)$ to be 1 if $a > 0$, -1 if $a < 0$ and 0 if $a = 0$. For any player $i \in \{1, \ldots, N\}$, let $\tilde{\gamma}_i^*(x)$ denote the conditional expectation of the product $D_i(\Sigma_{j \neq i} D_j)$ given x observed in the data. That is, $\tilde{\gamma}_i^*(x) \equiv \int_{p^l \in \mathcal{L}_x} p_i^l(x)\gamma_i^l(x) d\Lambda_x(p^l)$, where Λ_x denotes the equilibrium-selection mechanism in the DGP. Let $p_i^*(x)$ be the actual probability that i chooses 1 given x observed in the data (i.e. $p_i^*(x) = \int_{p^l \in \mathcal{L}_x} p_i^l(x) d\Lambda_x(p^l)$), and let $\gamma_i^*(x) \equiv \Sigma_{j \neq i} p_j^*(x)$. Let $\mathcal{L}_x^+$ denote the subset of $\mathcal{L}_x$ that occurs in the DGP with positive probability ($\mathcal{L}_x^+ \equiv \{p^l : \Lambda_x(p^l) > 0\}$). Multiple BNE exist in the DGP if $\mathcal{L}_x^+$ is not a singleton.

Proposition 1 *Suppose Assumption 1 holds. (i) For any given x, multiple BNE exist in the data-generating process if and only if $\tilde{\gamma}_i^*(x) \neq p_i^*(x)\gamma_i^*(x)$ at least for some i; (ii) For all i and x such that $\tilde{\gamma}_i^*(x) \neq p_i^*(x)\gamma_i^*(x)$,*

$$sign\left(\tilde{\gamma}_i^*(x) - p_i^*(x)\gamma_i^*(x)\right) = sign(\delta_i(x)) \quad (2)$$

Proof of Proposition 1. Under Assumption 1, D_i must be independent of $\Sigma_{j \neq i} D_j$ conditional on x in every single BNE p^l in $\mathcal{L}_x$.

(Sufficiency of (i)) Suppose there is a unique BNE in the data-generating process. That is, $\mathcal{L}_x^+$ is a singleton $\{p^l\}$. Then $p_i^*(x) = p_i^l(x)$, $\gamma_i^*(x) = \Sigma_{j \neq i} p_j^l(x)$ and $\tilde{\gamma}_i^*(x) = p_i^l(x)\Sigma_{j \neq i} p_j^l(x)$ for all i in state x. Hence $\tilde{\gamma}_i^*(x) = p_i^*(x)\gamma_i^*(x)$ for all i.

(Necessity of (i)) Suppose $\mathcal{L}_{x,\theta}^+$ is not a singleton in state x. Then there exists at least some i and $p^l, p^k \in \mathcal{L}_x^+$ such that $p_i^l \neq p_i^k$.Also note that for such a player i, $\delta_i(x)$ must necessarily be non-zero. By definition,

$$\begin{aligned} \Delta_i(x) &\equiv \tilde{\gamma}_i^*(x) - p_i^*(x)\gamma_i^*(x) \\ &= \int_{p^l \in \mathcal{L}_x^+} p_i^l(x)\gamma_i^l(x) d\Lambda_x - \int_{p^l \in \mathcal{L}_x^+} p_i^l(x) d\Lambda_x \int_{p^l \in \mathcal{L}_x^+} \gamma_i^l(x) d\Lambda_x \end{aligned} \quad (3)$$

Suppose $\delta_i(x) > 0$. The equilibrium characterization in (1) implies that there exists a strictly increasing function h_i such that $\gamma_i^l(x) = h_i(p_i^l(x)) \equiv \left(F_{\epsilon_i|X}^{-1}(p_i^l(x)) - u_i(x)\right)/\delta_i(x)$ for each single p^l in $\mathcal{L}_{x,\theta}$.[5] Thus for x given, (3) can be written as

$$\tilde{\gamma}_i^*(x) - p_i^*(x)\gamma_i^*(x) = \int_0^1 h_i(z) z d\tilde{\Lambda}_{i,x}(z) - \int_0^1 z d\tilde{\Lambda}_{i,x}(z) \int_0^1 h_i(z) d\tilde{\Lambda}_{i,x}(z)$$

where $z \equiv p_i^l(x)$ and $\tilde{\Lambda}_{i,x}$ is a distribution of $p_i^l(x)$ induced by the equilibrium selection mechanism Λ_x defined on $\mathcal{L}_x$. Thus (3) takes the simple form of the covariance of a random variable z and a strictly increasing function of itself:

$$\begin{aligned}
\text{cov}(Z, h_i(Z)) &= \mathbb{E}\left[(Z - \mathbb{E}(Z))(h_i(Z) - \mathbb{E}(h_i(Z)))\right] \\
&= \mathbb{E}\left[(Z - \mathbb{E}(Z))(h_i(Z) - h_i(\mathbb{E}(Z)))\right] + \mathbb{E}\left[(Z - \mathbb{E}(Z))(h_i(\mathbb{E}(Z)) - \mathbb{E}(h_i(Z)))\right] \\
&= \mathbb{E}\left[(Z - \mathbb{E}(Z))(h_i(Z) - h_i(\mathbb{E}(Z)))\right]
\end{aligned}$$

Because h_i is strictly increasing in $[0,1]$ for given x, we have $z_1 > z_2 \Rightarrow h_i(z_1) > h_i(z_2)$. Consequently, $(z - \mathbb{E}(Z))(h_i(z) - h_i(\mathbb{E}(Z))) > 0$ for any $z \neq \mathbb{E}(Z)$, and the covariance is strictly positive, provided the distribution $\tilde{\Lambda}_{i,x}$ is not degenerate on $\mathcal{L}_x^+$. Hence $\tilde{\gamma}_i^*(x) - p_i^*(x)\gamma_i^*(x) > 0$ if multiple BNE exist in the data-generating process in state x. The case with $\delta_i(x) < 0$ is proved by symmetric arguments. The proof of (ii) is already included in the proof of (i) above. ■

Part (i) of Proposition 1 can be exploited to devise a Wald Test for multiple BNE under any given x in the DGP. We describe the test and discuss its asymptotic properties in the appendix. Part (ii) of the proposition suggests the sign of $\delta_i(x)$ can be recovered from observed distributions provided i actively switches between multiple equilibrium strategies under x in DGP.

In some empirical contexts, players' actions may have heterogeneous impacts on each others' payoffs. Our arguments in Proposition 1 can be extended as long as econometricians know the role of these heterogeneities in strategic interactions. More specifically, we allow $U_{1i}(X, \epsilon_i) \equiv u_i(X) + \delta_i(X) f_i(X, D_{-i}) - \epsilon_i$, where $f_i(X, D_{-i})$ is a known function summarizing how individual actions affect interaction effects and $\delta_i(x)$ is a baseline effect whose sign is to be inferred.

[5] The form of h_i may depend on θ and x in general. We suppress this dependence for notational ease.

For a fixed x, any function $f_i(x, D_{-i})$ can take at most 2^{N-1} values corresponding to the possible D_{-i} vectors: $\{f_i(x,\pi) : \pi \in \{0,1\}^{N-1}\}$. We can then write $f_i(x, D_{-i}) = \sum_{\pi\in\{0,1\}^{N-1}} f_i(x,\pi)\prod_{j\neq i} 1\{D_j = \pi_j\} = \sum_{\pi\in\{0,1\}^{N-1}} f_i(x,\pi)\prod_{j\neq i} D_j^{\pi_j}(1-D_j)^{1-\pi_j}$ where π_j denotes the jth component of π and $1\{.\}$ is the indicator function. For example, if $N = 3$ and $f_i(x, D_{-i}) = \max_{j\neq i}(D_j)$ we have that $f_1(x, D_{-1}) = \max\{1,1\}D_2D_3 + \max\{1,0\}D_2(1-D_3) + \max\{0,1\}(1-D_2)D_3 + \max\{0,0\}(1-D_2)(1-D_3) = D_2(1-D_3) + D_3(1-D_2) + D_2D_3$ (and analogously for $i = 2, 3$). By Assumption 1, in a single equilibrium indexed by l,

$$\begin{aligned}
\mathbb{E}[f_i(X, (S_j^l(X,\epsilon_j))_{j\neq i})|X = x, \epsilon_i] &= \mathbb{E}[f_i(X, (S_j^l(X,\epsilon_j))_{j\neq i})|X = x] \\
&= \sum_{\pi\in\{0,1\}^{N-1}} f_i(x,\pi)P^l(D_{-i} = \pi|x) \\
&= \sum_{\pi\in\{0,1\}^{N-1}} \left[f_i(x,\pi)\prod_{j\neq i} p_j^l(x)^{\pi_j}(1-p_j^l(x))^{1-\pi_j}\right] \\
&\equiv \phi_{f_i}(x, p_{-i}^l(x))
\end{aligned}$$

where $p_j^l(x) \equiv \mathbb{E}(S_j^l(X,\epsilon_j)|X = x)$ as before, $p_{-i}^l(x) \equiv (p_j^l(x))_{j\neq i}$, and $P^l(\omega|x)$ denotes the probability that "the event ω happens conditional on x" as implied in the equilibrium p^l. Notice also that the mapping $\phi_{f_i} : \Omega_X \times [0,1]^{N-1} \to \mathbb{R}$ is a simple extension of $f_i : \Omega_X \times \{0,1\}^{N-1} \to \mathbb{R}$ to $\Omega_X \times [0,1]^{N-1}$. It is known as long as f_i is known. The equations characterizing a single equilibrium p^l in (1) now become:

$$p_i^l(x) = F_{\epsilon_i|X=x}\left(u_i(x) + \delta_i(x)\phi_{f_i}(x, p_{-i}^l(x))\right) \textit{ for all } i = 1, \ldots, N$$

Then the results in Proposition 1 now apply with $\gamma_i^l(x) \equiv \phi_{f_i}(x, p_{-i}^l(x))$. Note that, by the Law of Total Probability,

$$\begin{aligned}
\gamma_i^*(x) &\equiv \int_{p^l\in\mathcal{L}_x} \phi_{f_i}(x, p_{-i}^l(x))d\Lambda_x(p^l) = \int_{p^l\in\mathcal{L}_x} \sum_{\pi\in\{0,1\}^{N-1}} [f_i(x,\pi)P^l(D_{-i} = \pi|x)]d\Lambda_x(p^l) \\
&= \sum_{\pi\in\{0,1\}^{N-1}} [f_i(x,\pi)P^*(D_{-i} = \pi|x)] \qquad (4)
\end{aligned}$$

where $P^*(\omega|x)$ denotes the probablity that "ω occurs conditional on x" observed from the data. Furthermore,

$$\begin{aligned}\tilde{\gamma}_i^*(x) &\equiv \int_{p^l\in\mathcal{L}_x} p_i^l(x)\phi_{f_i}(x,p_{-i}^l(x))d\Lambda_x(p^l) \\ &= \int_{p^l\in\mathcal{L}_x}\sum_{\pi\in\{0,1\}^{N-1}}[f_i(x,\pi)P^l\{(D_{-i},D_i)=(\pi,1)|x\}]d\Lambda_x(p^l) \\ &= \sum_{\pi\in\{0,1\}^{N-1}}[f_i(x,\pi)P^*\{(D_{-i},D_i)=(\pi,1)|x\}] \qquad (5)\end{aligned}$$

where the first equality follows from $P^l\{(D_{-i},D_i) = (\pi,1)|x\} = p_i^l(x)\prod_{j\neq i}p_j^l(x)^{\pi_j}(1-p_j^l(x))^{1-\pi_j}$ under Assumption 1, and the second equilibrium follows from the Law of Total Probability and the definition of $P^*(\cdot|x)$. Hence $p_i^*(x)$, $\gamma_i^*(x)$ as defined in (4) and $\tilde{\gamma}_i^*(x)$ as defined in (5) can all be expressed in terms of observable distributions. Thus the sign of $\delta_i(x)$ is identified and multiple BNE can be detected as in Proposition 1.

3.2 Allowing for unique BNE

The result in part (ii) of Proposition 1 shows that the sign of interaction effects for i under x can be recovered provided that there exist multiple BNE at x in the DGP and that i follows different strategies across these equilibria. This result does not warrant the identification of $sign(\delta_i(x))$ for all (i,x), because there can exist players who employ the same strategies across all equilibria under x. This could happen when there is a unique BNE under state x. It could also occur if the game involves three or more players and, for some player i, all of the multiple BNE under x prescribe the same strategy. (That is, there is i with $p_i^l = p_i^*$ for all p_i^l in $\mathcal{L}_{x,\theta}^+$, so that $\tilde{\gamma}_i^*(x) = p_i^*(x)\gamma_i^*(x)$.) The following example illustrates this possibility.

Example 1. *(A player who follows the same strategy in multiple BNE)* Consider a simple 3-by-2 game with $N = 3$, where the identities of all three players are observable in data. Suppress the dependence on x for notational ease. Let $u_1 = 0.5$, $u_2 = u_3 = 0.3611$, $\delta_i = -1$

and $\epsilon_i \sim N(0.10, 0.25^2)$ for all i. Then there exist at least two distinct BNE:

$$
\begin{aligned}
p^a \text{ with } p_1^a &= 0.0611;\ p_2^a = 0.7756;\ p_3^a = 0.0107 \\
p^b \text{ with } p_1^b &= 0.0611;\ p_2^b = 0.0107;\ p_3^b = 0.7756
\end{aligned}
$$

In these two BNE (p^a and p^b), Player 1 chooses alternative 1 with the same probability in both BNE, while both 2 and 3 play strategies that imply different choice probabilities in equilibrium (i.e. $p_i^a \neq p_i^b$ for $i = 2, 3$). $\|$

This issue can be solved if, for the (i, x) considered, the signs or the magnitudes of the interaction effects are known to remain the same over a set of covariate realizations (for example because of parameter constancy or, more generally, exclusion restrictions). In such cases, the researcher can pool information from games with heterogeneous covariates to help identify the signs of interaction effects for such a (i, x). We consider these two scenarios for the rest of this subsection.

Aggregating data from games with the same sign of $\delta_i(x)$

Consider a simplified case where strategic interaction effects have the same sign for all $x \in \Omega_X$ for some i. Then $sign(\delta_i(\cdot))$ is identified if and only if the set of states where i uses multiple BNE strategies in the DGP has a positive measure under F_X. To see this, note that $\delta_i(x) > (<)\ 0$ if $\tilde{\gamma}_i^*(x) - p_i^*(x)\gamma_i^*(x) > (<)\ 0$. Furthermore, if $\delta_i(x) > (<)\ 0$ and multiple equilibria are played in the DGP under x, then $\tilde{\gamma}_i^*(x) - p_i^*(x)\gamma_i^*(x) > (<)\ 0$. It then follows that if the set of x under which i adopts multiple BNE strategies occurs with positive probability, then the sign of $\mathbb{E}[\tilde{\gamma}_i^*(X) - p_i^*(X)\gamma_i^*(X)]$ is the same as the sign of $\delta_i(\cdot)$. On the other hand, if i sticks to a single BNE strategy for (F_X-almost) every x, then $\tilde{\gamma}_i^*(x) = p_i^*(x)\gamma_i^*(x)$ F_X-a.e. and $\mathbb{E}[\tilde{\gamma}_i^*(X) - p_i^*(X)\gamma_i^*(X)] = 0$. The following corollary formalizes and generalizes this idea.

Corollary 1 *Suppose Assumption 1 holds and there is a known set ω_i such that $sign(\delta_i(.))$ remains the same for all $x \in \omega_i$. Then (i) $sign(\delta_i(.))$ is recovered on ω_i as the sign of $\mathbb{E}[\tilde{\gamma}_i^*(X) - p_i^*(X)\gamma_i^*(X)|X \in \omega_i]$ if*

$$\Pr\{x \in \omega_i : \text{ } i \textit{ follows multiple BNE strategies at } x\} > 0 \tag{6}$$

and (ii) the condition in (6) holds if and only if

$$\mathbb{E}[\tilde{\gamma}_i^*(X)|X \in \omega_i] \neq \mathbb{E}[p_i^*(X)\gamma_i^*(X)|X \in \omega_i].$$

Corollary 1 shows that $sign(\delta_i(x))$ can be identified even when there is a unique BNE at x, as long as i employs multiple BNE strategies with positive probability over a set of x' with $sign(\delta_i(x)) = sign(\delta_i(x'))$. The corollary is a straightforward consequence of Proposition 1.

Aggregating data from games with the same size of $\delta_i(x)$

So far identification of $\delta_i(x)$ has relied on existence of multiple BNE. For the rest of this section, we consider a DGP where the BNE adopted at each x may be unique. We show that $sign(\delta_i(x))$ can still be recovered in this case if an exclusion restriction holds. This strategy is also invoked in similar contexts in the literature. To understand this exclusion restriction, consider a game involving N firms which make simultaneous entry or exit decisions. The vector of states X include a subvector $\tilde{X}_0$ consisting of market- or sector-wide factors that affect the demand for firm products. X also includes mutually exclusive subvectors $(\tilde{X}_i)_{i\leq N}$ with $\tilde{X}_i$ capturing firm-specific factors that only affect the profits for Firm i but none of its rivals. For example, $\tilde{X}_i$ may include labor costs or local regulations pertaining to the geographic location of i. The vector of private information $(\epsilon_i)_{i\leq N}$ may well capture all other firm-specific profit factors (such as idiosyncratic costs) that are unobservable to opponents and econometricians. If rivals' idiosyncratic factors (such as labor costs) have no bearing on Firm i's profits in addition to $\tilde{X}_0$ and $\tilde{X}_i$, then $F_{\epsilon_i|X} = F_{\epsilon_i|X_i}$ where $X_i = (\tilde{X}_0, \tilde{X}_i)$. For each

(i,x), we refer to the set $\Upsilon_i(x_i) \equiv \{x' : x'_i = x_i\}$ as the *equivalence class* for i at x. We state the exclusion restriction assumption as follows:

Assumption 2 *For all i, there exists a strict subvector of X (denoted X_i) such that $u_i(x) = u_i(x_i)$, $\delta_i(x) = \delta_i(x_i)$, $F_{\epsilon_i|X=x} = F_{\epsilon_i|X_i=x_i}$ for all x.*

The main idea for identifying $sign(\delta_i(x))$ (even when i only has a unique BNE strategy at each realization x) is based on three observations: (a) Player i can adopt different BNE strategies across games with states in the equivalence class (as long as $u_j(x'), \delta_j(x'), F_{\epsilon_j|X=x'}$ vary over $\Upsilon_i(x_i)$); (b) By assumption, the equilibrium conditions relating player i's strategies and those of the rivals' in any single BNE characterized by (1) must take the *same* form for all x' in $\Upsilon_i(x_i)$; and (c) The way opponents' choice probabilities affect i's choice probability across different x' in $\Upsilon_i(x_i)$ is only determined by the sign of the strategic interaction effect for i, which is the same for all x' in $\Upsilon_i(x_i)$ under the exclusion restriction in Assumption 2. Consequently, we can use an argument similar to that in Section 3.1 to identify $sign(\delta_i(x))$. If in response to her opponents' equilibrium strategies, i is induced to adopt different BNE strategies across games with different states in the equivalence class for x, then the sign of the correlation between actions by i and competitors across these games identifies $sign(\delta_i(x))$ just as in our previous analysis.

Let $\Lambda^*_{x_i}$ be the probability distribution over equilibrium choice probability profiles in the equivalence class for x_i. It is obtained by integrating the equilibrium selection mechanism Λ_x across the states in $\Upsilon_i(x_i)$ with respect to conditional distribution $F_{X|X\in\Upsilon_i(x_i)}$. That is, for any $A \subseteq [0,1]^N$,

$$\Lambda^*_{x_i}(A) \equiv \int_{\{x\ :\ \mathcal{L}^+_x \cap A \neq \varnothing\}} \Lambda_x(\mathcal{L}^+_x \cap A) dF_{X|X\in\Upsilon_i(x_i)}(x)$$

where Λ_x denotes the equilibrium selection probabilities defined in Section 2. It is easy to verify that $\Lambda^*_{x_i}$ is a well-defined distribution. Let its support be denoted by $\mathcal{L}^*_{x_i}$.

The key condition for identifying $\delta_i(x)$ is that player i adopts varying strategies across BNE in different games whose states belong to the equivalence class for i at x. Formally, the distribution $\Lambda^*_{x_i}$ is non-degenerate in i's dimension if $\nexists t \in [0,1]$ such that the support $\mathcal{L}^*_{x_i} \subseteq \{p \in [0,1]^N : p_i = t\}$. We give a simple illustration of Assumption 2 and the non-degeneracy condition in Design 2 of the Monte Carlo section. We also this assumption in greater detail following Proposition 2. Let g index independent games observed in data, and let $D_{i,g}$ denote the decision made by i in game g. Define

$$\Psi_i(x_i) \equiv \mathbb{E}[D_{i,g}\left(\Sigma_{j\neq i}D_{j,g}\right)|X_g \in \Upsilon_i(x_i)] - \mathbb{E}[D_{i,g}|X_g \in \Upsilon_i(x)]\mathbb{E}[\Sigma_{j\neq i}D_{j,g}|X_g \in \Upsilon_i(x_i)]$$

Proposition 2 *Suppose Assumptions 1 and 2 hold. Then (i) at any x, $sign(\delta_i(x)) = sign(\Psi_i(x_i))$ for all i if $\Lambda^*_{x_i}$ is non-degenerate in i's dimension; and (ii) $\Lambda^*_{x_i}$ is non-degenerate in i's dimension if and only if $\Psi_i(x_i) \neq 0$.*

Proof of Proposition 2. Consider any pair of (i, x) such that $\Lambda^*_{x_i}$ is not degenerate. The equations in (1) and Assumption 2 imply that there exists a function h_i such that $\gamma^l_i(z) = h_i(p^l_i(z))$ for all $z \in \Upsilon_i(x_i)$ and $p^l \in \mathcal{L}^+_x$, where $h_i(\cdot) \equiv \left(F^{-1}_{\epsilon_i|x}(\cdot) - u_i(x)\right)/\delta_i(x)$. The function h_i summarizes the interdependence between i's BNE strategies and those for $j \neq i$. If $\Lambda^*_{x_i}$ is non-degenerate in i's dimension, then :

$$\begin{aligned}\Psi_i(x_i) &= \int_{p\in\mathcal{L}^*_{x_i}} \mathbb{E}[D_i\left(\Sigma_{j\neq i}D_j\right)|p, X \in \Upsilon_i(x_i)]d\Lambda^*_{x_i} \\ &\quad - \left(\int_{p\in\mathcal{L}^*_{x_i}} \mathbb{E}[D_i|p, X \in \Upsilon_i(x_i)]d\Lambda^*_{x_i}\right)\cdot\left(\int_{p\in\mathcal{L}^*_{x_i}} \mathbb{E}[\Sigma_{j\neq i}D_j|p, X \in \Upsilon_i(x_i)]d\Lambda^*_{x_i}\right) \\ &= \int_{p\in\mathcal{L}^*_x} p_i\left(\Sigma_{j\neq i}p_j\right)d\Lambda^*_{x_i} - \int_{p\in\mathcal{L}^*_x} p_i d\Lambda^*_{x_i}\int_{p\in\mathcal{L}^*_x}\left(\Sigma_{j\neq i}p_j\right)d\Lambda^*_{x_i} \qquad (7)\end{aligned}$$

where $p \in [0,1]^N$ denotes a generic characterization of BNE on the support $L^*_{x_i}$. The first equality follows from the definition of $\Lambda^*_{x_i}$ and the second from independence between D_i and $(D_j)_{j\neq i}$ conditional on the equilibrium played and on states being in the equivalence class. Because $h_i(\cdot)$ is the same for all $x \in \Upsilon_i(x_i)$ due to Assumption 2, (7) can be written as:

$$\Psi_i(x_i) = \int_0^1 p_i h_i(p_i)d\tilde{\Lambda}^*_{x_i} - \int_0^1 p_i d\tilde{\Lambda}^*_{x_i}\int_0^1 h_i(p_i)d\tilde{\Lambda}^*_{x_i} \qquad (8)$$

where $\tilde{\Lambda}^*_{x_i}$ is the marginal distribution of p_i according to the joint distribution $\Lambda^*_{x_i}$. Finally, note $h_i(\cdot)$ is increasing (or decreasing) over $[0, N-1]$ if $\delta_i(x) > 0$ (or < 0) for $x \in \Upsilon_i(x_i)$.

Hence the same argument as in Proposition 1 shows that for all (i, x), $\Psi_i(x_i) > 0$ (or < 0) if $\delta_i(x) > 0$ (or < 0) and $\Lambda^*_{x_i}$ is non-degenerate in i's dimension. It also follows immediately from (8) that if $\Lambda^*_{x_i}$ is degenerate in i's dimension, then $\Psi_i(x_i) = 0$. ■

That the distribution $\Lambda^*_{x_i}$ is non-degenerate in i's dimension is a weak restriction given Assumption 2. For this to hold, it is necessary that $\delta_i(x) \neq 0$ and $(u_j, \delta_j, F_{\epsilon_j|X})$ for $j \neq i$ vary over states in the equivalence class for i. The non-degeneracy can fail in cases such as when player i does not interact with rivals at all at x ($\delta_i(x) = 0$).[6] Part (ii) of Proposition 2 suggests an immediate test for the non-degeneracy condition using observed distributions of states and actions. The example below shows how the non-degeneracy condition can hold for all (i, x) under fairly intuitive restrictions.

Example 2. *(Non-degeneracy for all i, x)* Consider a 2-by-2 entry or exit game with incomplete information between Firm 1 and 2 with state vector X which can be partitioned as $(\tilde{X}_0, \tilde{X}_1, \tilde{X}_2)$, where $\tilde{X}_0$ are market-level factors that affect profitability of the firms and $\tilde{X}_1$ and $\tilde{X}_2$ are firm-level characteristics for Firm 1 and 2, respectively. Suppose $\delta_i(x) \neq 0$ for all i, x and Assumptions 1 and 2 hold with $X_1 \equiv (\tilde{X}_0, \tilde{X}_1)$ and $X_2 \equiv (\tilde{X}_0, \tilde{X}_2)$. Assume further that the interaction effects and the distribution private information only depend on market level factors $\tilde{X}_0$ ($\delta_i(x_i) = \delta_i(\tilde{x}_0)$ and $F_{\epsilon_i|X_i=x_i} = F_{\epsilon_i|\tilde{X}_0=\tilde{x}_0}$ for both i and all x_i). Then for $i = 1, 2$, the probability of entering (choosing action 1) in a BNE is given by:

$$p_i(x) = F_{\epsilon_i|\tilde{x}_0}(u_i(x_i) + \delta_i(\tilde{x}_0)p_{3-i}(x))$$

Assume for any $x_1 = (\tilde{x}_0, \tilde{x}_1)$, there exists a set of $\tilde{x}_2$ (denoted ω_2) that occur with positive probability and leads to different baseline profits. That is, $\Pr\{\tilde{X}_1 \in \omega_2 | x_1\} > 0$ and $u_2(\tilde{x}_0, \tilde{x}_2) \neq u_2(\tilde{x}_0, \tilde{x}'_2)$ for all $\tilde{x}_2 \neq \tilde{x}'_2$ in ω_2. It then follows that for any pair of states $x \equiv (\tilde{x}_0, \tilde{x}_1, \tilde{x}_2)$ and $x' \equiv (\tilde{x}_0, \tilde{x}_1, \tilde{x}'_2)$, $\Lambda^*_{x_1}$ is non-degenerate in i's dimension for any x_1.[7]

[6] When $\delta_i(x) \neq 0$, this condition can also fail if best responses for $j \neq i$ change over the equivalence class for i at x in very peculiar ways so that the solution for $p_i(\cdot)$ in (1) for $x \in \Psi_i(x_i)$ remains the same.

[7] To see this, note $F_{\epsilon_1|\tilde{X}_0=\tilde{x}_0}(u_1(\tilde{x}_0, \tilde{x}_1) + \delta_1(\tilde{x}_0)t)$ as a function of t over $[0, 1]$ remains the same for x and x', while $F_{\epsilon_2|\tilde{X}_0=\tilde{x}_0}(u_2(\tilde{x}_0, \tilde{x}_2) + \delta_2(\tilde{x}_0)t) \neq F_{\epsilon_2|\tilde{X}_0=\tilde{x}_0}(u_2(\tilde{x}_0, \tilde{x}'_2) + \delta_2(\tilde{x}_0)t)$ for all $t \in [0, 1]$.

Swapping 1 with 2 and repeating the arguments above shows $\Lambda^*_{x_2}$ can be non-degenerate in 2's dimension for any x_2. $\|$

We conclude our discussion by noting that the exclusion restriction in Assumption 2 is stronger than necessary for identifying $sign(\delta_i(x))$. In fact the preceding arguments can be extended easily to accommodate general forms of equivalence classes $\{x' \in \Omega_X : u_i(x') = u_i(x)\ ,\ \delta_i(x') = \delta_i(x) \text{ and } F_{\epsilon_i|x} = F_{\epsilon_i|x'}\}$. In particular, if no variables are excluded for individual i and the equivalence class for i at x is a singleton, then the non-degeneracy of Λ^*_x on i's dimension will amount to the existence of multiple BNE strategies at state x.

4 Testing Multiple BNE and Inferring Interaction Signs

The test for multiple equilibria is of practical importance in structural empirical research. When the equilibrium conditional choice probabilities are the same for all players in a game, the average choice in each game is an unbiased estimator for the conditional choice probabilities *within a particular equilibrium* (see, for example, Brock and Durlauf (2007), p.58). However, even when all players have identical payoff functions ($u_i(\cdot)$ and $\delta_i(\cdot)$) and private information distributions ($F_{\epsilon_i|X}$) though, asymmetric Bayesian Nash equilibria with different conditional choice probabilities across players may arise. This will happen for instance when the (common) $\delta(\cdot)$ is negative. When the equilibrium conditional choice probabilities differ across players and/or number of players in each game is small (as is typically the case in the empirical games literature), the conditional choice probabilities will not be reliably estimated within individual games. It is then necessary to pool data across games in which the same equilibrium is played so as to estimate the choice probabilities using more data. In this case, testing for multiple equilibria is of interest in its own right.

Besides, most of the known methods for semi-parametric estimation of incomplete information games (without explicitly specifying an equilibrium selection rule) have relied on the existence of a single equilibrium in the data (e.g. Aradillas-Lopez (2010), Bajari, Hong, Krainer, and Nekipelov (2010) and Tang (2010)).[8] Hence it is *imperative* to devise a formal test for the assumption of unique equilibrium in the data-generating process.

We focus on an empirical context where researchers observe states and decisions from a large cross-section of independent games (indexed by $g = 1, \ldots, G$) drawn from the same DGP characterized by $(u_i, \delta_i)_{i \leq N}, F_{\epsilon|X}$. Consider the null hypothesis that a "unique BNE exists in the DGP under state x ". By Proposition 1, the null of a unique BNE in the DGP is equivalently formulated as:

$$H_0 \quad : \quad \Delta_i(x) = 0 \ \forall i \leq N \tag{9}$$

We confront this null hypothesis with the alternative that:

$$H_1 \quad : \quad \exists i \ s.t. \ \Delta_i(x) \neq 0 \tag{10}$$

where

$$\Delta_i(x) \equiv \tilde{\gamma}_i^*(x) - p_i^*(x)\gamma_i^*(x) = \sum\nolimits_{j \neq i} \left\{ \mathbb{E}[D_i D_j|x] - E(D_i|x)E(D_j|x) \right\}.$$

It follows from Section 3 that $\Delta_i(x) \neq 0$ if and only if i adopts multiple strategies with positive probability at x. In the appendix we propose a simple Wald Test that can be used to test the joint null in (9) at x. We also note that the parameter $\Delta_i(x)$ can be easily adapted to accommodate general (known) $f_i(x, D_{-i})$ as indicated previously.

A failure to reject the null of unique equilibrium in the DGP suggests the equilibrium conditions in (1) can be used for estimation. It is then possible to invoke additional assumptions on $u, \delta, F_{\epsilon|x}$ to identify the model structure. Examples of such restrictions include the index utilities and statistical or median independence of private information or even the knowledge of F. In implementation, sampling errors from such a pre-test for unique equilibrium should ideally be accounted for in deriving asymptotic properties. If the joint null of unique BNE is rejected, then finding out which of the N single nulls in (1) are responsible

[8]For an illustration of how misspecification of the equilibrium selection rule can affect inference in a complete information game with a small number of players, see Honoré and de Paula (2010).

for the rejection can help with robust estimation. We further motivate and address this question using a multiple-testing procedure in Section 4.1 below.

Finally, note that with $N = 2$ multiple BNE exist at x only if signs of individual interaction effects are the same for both players. In this case, both players adopt strategies that imply distinct conditional choice probabilities across these BNE. Testing for multiple equilibria and inference of signs of interaction effects can be done by testing sample correlation of actions between the two players (given X). In this case, inference of multiple BNE and signs of interaction effects will be based on a scalar statistic $T_G \equiv T_{G,1} = T_{G,2}$ as defined later in this section.

4.1 Inference of Players With Multiple Equilibrium Strategies

With $N \geq 3$, while a subset of the players may employ different strategies across multiple BNE in the DGP, others might stick to the same strategy in all games observed in the data (see Example 1). Finding out the set of players who adopt multiple strategies has important implications for identifying and estimating players' payoffs. Semi-parametric estimation of Bayesian games typically refrains from parametric restrictions on primitives or the equilibrium selection mechanism at the cost of assuming that there is only a unique DGP for all x in the data. The applicability of these robust estimation approaches hinges on this single equilibrium assumption.[9] While a simple test of the joint null (9) using Wald statistics helps detect existence of multiple BNE, it does not specify any rules for deciding which players employ different strategies across multiple BNE.

Since we would like to detect which players employ different strategies across BNE and make inference on those players' interaction effects signs, we resort to the statistical literature on multiple comparisons (for a recent survey, see Lehmann and Romano (2005),

[9]It should be noted that "social interaction" models do not rely on this assumption but require the number of agents in each game to be large so that *within (symmetric) equilibrium* choice probabilities can be consistently estimated from average choices in each game.

Chapter 9). This literature considers decision strategies that aggregate the tests for the individual hypotheses corresponding to each i given x:

$$H_i^0 \quad : \quad \Delta_i(x) = 0$$
$$H_i^1 \quad : \quad \Delta_i(x) \neq 0$$

Given individual test statistics for each of the $i \leq N$ hypotheses, our objective is to define a decision rule that controls the *family-wise error rate*, or the probability of rejecting at least one of the true null hypotheses. That is:

$$\text{FWE}_P = \text{Prob}_P\{\text{Reject at least one } H_i^0 : \Delta_i(x) = 0 \text{ where } i \in \mathbf{I}_0(P)\}$$

where the subscript P indicates the DGP and $\mathbf{I}_0(P) \subset \{1, \ldots, N\}$ is the set of indices i of true null hypotheses under P. A multiple testing procedure asymptotically controls the FWE_P at α if $\limsup_{G \to +\infty} \text{FWE}_P \leq \alpha$ for any P.

We focus on a finite support Ω_X and we suppress x for notational ease when there is no ambiguity. Sample analogs of expectations conditional on x are simply calculated as the sample averages across games with $X = x$. Whereas this is easily done when Ω_X is discrete, a sample analog for a continuous X would involve the aggregation of realizations at "nearby" observations via nonparametric techniques (e.g. kernel methods). Since covariates may induce a different number of equilibria, in small samples the inference for a particular realization in Ω_X may be contaminated by the uniqueness or multiplicity of solutions at neighboring realizations. Note nevertheless that the identification arguments do not require that Ω_X have finite support. A thorough analysis of this inference problem under continuous covariates is beyond the scope of this paper.

We focus on the case with $N \geq 3$. For any subset $I \subset \{1, \ldots, N\}$, let $D_{I,g} \equiv \Pi_{i \in I} D_{i,g}$, $\mu_I \equiv E(D_{I,g} 1(X_g \in \{x\}))$ and $\mu_0 \equiv \Pr(X_g \in \{x\})$.[10] In addition, $\mu(\{x\})$ denotes a $\tilde{N} \equiv \left(N + \binom{N}{2} + 1\right)$-vector consisting of $\mu_0(\{x\}), \mu_i(\{x\})$ and $\mu_{ij}(\{x\})$ for all individual

[10] If $\mu_0(\{x\}) = 1$, an unconditional version of our procedure can be easily derived.

i and all pairs $i \neq j$. For example, with $N = 3$ (and omitting the argument $\{x\}$), $\mu \equiv (\mu_0, \mu_1, \mu_2, \mu_3, \mu_{12}, \mu_{13}, \mu_{23})'$. Define:

$$\begin{aligned} \hat{\mu}_i(\{x\}) &\equiv (G)^{-1} \sum\nolimits_g D_{i,g} 1(X_g \in \{x\}) \ ; \ \hat{\mu}_{ij}(\{x\}) \equiv (G)^{-1} \sum\nolimits_g D_{ij,g} 1(X_g \in \{x\}) \\ \hat{\mu}_0(\{x\}) &\equiv (G)^{-1} \sum\nolimits_g 1(X_g \in \{x\}) \ ; \ \hat{\mu}_G(\{x\}) \equiv (\hat{\mu}_0(\{x\}), \ldots, \hat{\mu}_i(\{x\}), \ldots, \hat{\mu}_{ij}(\{x\}), \ldots)' \end{aligned}$$

where $\hat{\mu}_G$ is the vector of sample analogs for μ. By the multivariate central limit theorem, $G^{1/2}(\overline{\mu}_G(\{x\}) - \mu(\{x\})) \xrightarrow{d} N(\mathbf{0}_{\tilde{N}}, \mathbf{\Sigma}(\{x\}))$ as $G \to \infty$, where $\mathbf{0}_{\tilde{N}}$ is a $\tilde{N}$-vector of zeros and Σ is the corresponding variance-covariance matrix. Define $\mathbf{T}_G(\{x\})$ to be a N-vector with its i-th coordinate being:

$$T_{G,i}(\{x\}) = \hat{\Delta}_i(x) \equiv \sum\nolimits_{j \neq i} \left(\frac{\hat{\mu}_{ij}(\{x\})}{\hat{\mu}_0(\{x\})} - \frac{\hat{\mu}_i(\{x\})\hat{\mu}_j(\{x\})}{(\hat{\mu}_0(\{x\}))^2} \right).$$

By the Delta Method, we obtain that

$$G^{1/2}(\mathbf{T}_G(\{x\}) - \mathbf{\Delta}(x)) \xrightarrow{d} N(\mathbf{0}_N, \mathbf{V}(\{x\})\mathbf{\Sigma}(\{x\})\mathbf{V}(\{x\})') \text{ as } G \to \infty$$

where $\mathbf{\Delta}(x) \equiv (\Delta_i(x))_{i=1}^N$. The Jacobian $\mathbf{V}(\{x\})$ is a N-by-$\tilde{N}$ matrix, with its i-th row $V_i(\{x\})$ defined by the following table (where $\mu_{(m)}(\{x\}), V_{i,(m)}(\{x\})$ denote the m-th coordinates of two $\tilde{N}$-vectors $\mu(\{x\})$ and $V_i(\{x\})$ respectively, and $j, k \neq i$),

$\mu_{(m)}(\{x\})$	$V_{i,(m)}(\{x\})$
$\mu_0(\{x\})$:	$\sum_{j \neq i}(-\frac{\mu_{ij}(\{x\})}{\mu_0(\{x\})^2} + \frac{2\mu_i(\{x\})\mu_j(\{x\})}{\mu_0(\{x\})^3})$
$\mu_i(\{x\})$:	$-\sum_{j \neq i} \frac{\mu_j(\{x\})}{\mu_0(\{x\})^2}$
$\mu_j(\{x\})$:	$-\frac{\mu_i(\{x\})}{\mu_0(\{x\})^2}$
$\mu_{ij}(\{x\})$ or $\mu_{ji}(\{x\})$:	$\frac{1}{\mu_0(\{x\})}$
$\mu_{jk}(\{x\})$:	0

We can estimate $\mathbf{\Sigma}(\{x\}), \mathbf{V}(\{x\})$ consistently by replacing $\mu_0(\{x\}), \mu_I(\{x\})$ with the sample analogs described above. For the remainder of this subsection, we omit the argument $(\{x\})$ for notational ease.

Well-known methods that asymptotically control for the family-wise error rate include the Bonferroni and the Holm's method. Both methods can be described in terms of the p-values for each of the individual hypotheses (indexed by i) above. We denote these p-values by $\hat{p}_{G,i}$. The Bonferroni method at level α rejects i if $\hat{p}_{G,i} \leq \alpha/N$. The Holm's procedure, which is less conservative than the Bonferroni method, follows a stepwise strategy. (For notational convenience, we suppress the dependence of the hypotheses and test statistics on x.) The Holm's procedure starts by ordering the p-values in ascending order: $\hat{p}_{G,(1)} \leq \hat{p}_{G,(2)} \leq \cdots \leq \hat{p}_{G,(N)}$. Let $H^0_{j_k} : \Delta_{j_k} = 0$ denote the single hypothesis corresponding to the k-th smallest p-value (i.e. $\hat{p}_{G,j_k} = \hat{p}_{G,(k)}$). Holm's stepwise method proceeds as follows. In the first step, compare $\hat{p}_{G,(1)}$ with α/N. If $\hat{p}_{G,(1)} \geq \alpha/N$, then accept all individual hypotheses and the procedure ends. Otherwise, reject the individual null hypothesis $H^0_{j_1} : \Delta_{j_1} = 0$ and move on to the second step. In the second step, the remaining $N-1$ hypotheses are all accepted if $\hat{p}_{G,(2)} \geq \alpha/(N-1)$. Otherwise reject $H^0_{j_2} : \Delta_{j_2} = 0$ and continue to the next step. More generally, compare $\hat{p}_{G,(k)}$ with $\alpha/(N-k+1)$ in the k-th step. Accept all remaining $N-(k-1)$ hypotheses if $\hat{p}_{G,(k)} \geq \alpha/(N-k+1)$. Otherwise, reject $H^0_{j_k}$ and move on to the next step. Continue doing so until all remaining hypotheses are accepted, or all hypotheses are rejected one by one in N steps.

Though less conservative than the Bonferroni method, the Holm's procedure can still be improved upon if one takes into account the dependence between individual test statistics. To achieve this, we follow recent contributions by van der Laan, Dudoit, and Pollard (2004) and Romano and Wolf (2005).[11] Ordering the test statistics in descending order, we let $T_{G,(1)} \geq T_{G,(2)} \geq \cdots \geq T_{G,(N)}$. In the k-th step, a critical level c_k is obtained and those hypotheses with $T_{G,\cdot} \geq c_k$ are rejected. Let R_k be the number of hypotheses rejected after the first $k-1$ steps (i.e. the number of hypotheses rejected at the beginning of the k-th step). As before, let $H^0_{i_k}$ denote the hypothesis whose test statistic is the k-th largest (i.e.

[11]The following description closely follows the presentation in Romano and Wolf (2005). For similar strategies controlling generalizations of the family-wise error rate, see Romano and Shaikh (2006). A recent application of such generalizations is Moon and Perron (2009).

$T_{G,i_k} = T_{G,(k)}$). Ideally, we want to obtain c_1 such that:

$$c_1 \equiv c_1(1-\alpha, P) = \inf\left\{y : \text{Prob}_P\left\{\max_{1\le j\le N} T_{G,(j)} - \Delta_{i_j} \le y\right\} \ge 1-\alpha\right\}$$

where all statements are implicitly conditional on $X = x$. Subsequently, c_k is defined as

$$c_k \equiv c_k(1-\alpha, P) = \inf\left\{y : \text{Prob}_P\left\{\max_{R_k+1\le j\le N} T_{G,(j)} - \Delta_{i_j} \le y\right\} \ge 1-\alpha\right\}$$

(also conditional on $X = x$). As pointed out in the references cited, because P is unknown in practice, we replace P by an estimate $\hat{P}_G$ and define

$$\hat{c}_k \equiv c_k(1-\alpha, \hat{P}_G) = \inf\left\{y : \text{Prob}_{\hat{P}_G}\left\{\max_{R_k+1\le j\le N} T^*_{G,(j)} - \Delta^*_{i_j} \le y\right\} \ge 1-\alpha\right\} \tag{11}$$

where we follow Romano and Wolf (2005) and use $T^*_{G,(j)}$ and $\Delta^*_{i_j}$ to highlight that the sampling distribution of the test statistics is under $\hat{P}_G$ (not P). The stepwise multiple testing procedure from Romano and Wolf (2005) can be summarized by algorithm A1 in the Appendix. In addition to estimating $\hat{c}_k$ via bootstrap, we also consider an alternative approach that uses the fact that the test statistics have a normal limiting distribution with a consistently estimable variance-covariance matrix.[12] We summarize the two approaches for estimating $\hat{c}_k$ in two algorithms A2 and A3 in the Appendix.

We can also use a studentized version of the multiple testing method as recommended in Romano and Wolf (2005). Let $\hat{\sigma}_{G,k}$ denote the estimates for the standard deviation of the test statistic $T_{G,k}$. To do so, we need an analogue of (11):

$$\hat{d}_k \equiv d_k(1-\alpha, \hat{P}_G) \equiv \inf\left\{y : \text{Prob}_{\hat{P}_G}\left\{\max_{R_k+1\le j\le N} (T^*_{G,(j)} - T_{G,i_j})/\hat{\sigma}^*_{G,i_j} \le y\right\} \ge 1-\alpha\right\}$$

where $\hat{\sigma}^*_{G,i}$ are the estimates for standard deviations of $T_{G,i}$ computed from bootstrap samples. A description of the procedure for the studentized statistic is presented in algorithm A.4 in the Appendix.

For a parametric model with state-independent interaction effects, Sweeting (2009) proposed two procedures to check for multiple symmetric equilibria in the data. The first

[12] See footnote 21 in Romano and Wolf (2005).

is based on calculating the percentage of pairs of players whose actions are correlated. The other is based on testing significance of equilibrium selection probabilities using Maximum Likelihood estimates with the number of BNE considered equal to two. Hence the second procedure is a test of the null of unique BNE against the alternative of two BNE in the DGP. In comparison, we develop stronger and new results by extending this intuition in a more general context. The most important distinction is that our test can be applied in cases where individual-specific interaction effects may depend on the states in unrestricted ways, and asymmetric equilibria may arise due to heterogeneities in players' payoffs. Furthermore, our test addresses several additional subtle issues. First, our test for multiple BNE in the data is based on testing whether each individual's action is correlated with an *aggregate* measure of competitors' actions. Therefore, our test has power under alternatives in which multiple BNE exist in the data with only a very small number of players switching strategies across the multiple equilibria. Second, our approach does not require knowledge of the number of equilibria in the alternative. Third, we apply a multiple testing procedure proposed by Romano and Wolf (2005) to test the joint null hypothesis that the equilibrium in the data is unique. And if the joint null is rejected, the procedure infers the exact identities of players who have switched between strategies in the data. Last, our test can be extended to allow for correlation private signals if researchers know a priori the groups/clusters of observed games within which the same equilibrium is played. (See Section 5.)

In Section 6, we report the performance of three tests based on stepwise multiple testing procedures: (a) the non-studentized test with $\hat{c}_k$ computed from parametric simulations; (b) the non-studentized test with $\hat{c}_k$ computed via bootstrap; and (c) the studentized test with $\hat{d}_k$ computed via bootstrap. Because our setting corresponds to the smooth function model with i.i.d. data (Scenario 3.1 in Romano and Wolf (2005)), both strategies yield consistent tests that asymptotically control the family-wise error rate at level α. This would obtain from a slight modification in Theorem 3.1 in Romano and Wolf (2005) to accommodate two-sided hypotheses as indicated in Section 5 of that paper.

4.2 Inference on Signs of Interaction Effects

This section proposes a simple test for the sign of interaction effects for a player i in a given state x. It relies on the characterization in Proposition 2 and will hold when x induces multiple equilibria and choice probabilities vary across equilibria or when there are excluded regressors as discussed in Section 3. We focus on the simple case with discrete X where any x in the support can happen with strictly positive probabilities. For any i, x, define

$$T_{G,i}(\Upsilon_i(x)) = \hat{\Delta}_i(x) \equiv \sum\nolimits_{j \neq i} \left(\frac{\hat{\mu}_{ij}(\Upsilon_i(x))}{\hat{\mu}_0(\Upsilon_i(x))} - \frac{\hat{\mu}_i(\Upsilon_i(x))\hat{\mu}_j(\Upsilon_i(x))}{(\hat{\mu}_0(\Upsilon_i(x)))^2} \right).$$

which is analogous to the statistic defined in the previous subsection, but with $1(X \in \Upsilon_i(x))$ in place of $1(X \in \{x\})$ when defining $\hat{\mu}_m$. This statistic is an estimator for Ψ_i introduced in Proposition 2. When $\Upsilon_i(x) = \{x\}$, this statistic coincides with the statistic introduced in subsection 4.1. For notational ease, we drop the subscript i, x from the estimators when there is no ambiguity. Using the Delta Method and Slutsky's Theorem it is straightforward to verify that

$$\left(\hat{\mathbf{V}}(\Upsilon_i(x))\hat{\mathbf{\Sigma}}(\Upsilon_i(x))\hat{\mathbf{V}}(\Upsilon_i(x))'/G\right)^{-1/2} (T_{G,i}(\Upsilon_i(x)) - \Psi_i(x)) \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as } G \to \infty$$

where $\hat{\mathbf{V}}(\Upsilon_i(x))$ and $\hat{\mathbf{\Sigma}}(\Upsilon_i(x))$ are estimators for $\mathbf{V}(\Upsilon_i(x))$ and $\mathbf{\Sigma}(\Upsilon_i(x))$, which themselves are defined analogously to the discussion in subsection 4.1. Testing the existence of multiple equilibria in the data *and* the sign of $\delta_i(x)$ amounts to testing the following three hypotheses:

$$H_+ : \Psi_i(x) > 0 \;;\; H_0 : \Psi_i(x) = 0 \;,\; H_- : \Psi_i(x) < 0.$$

Rejection of H_0 in favor of H_+ is indicative of multiple equilibria and a positive sign for $\delta_i(x)$. Analogously, rejection of H_0 in favor of H_- is indicative of multiple equilibria and a negative sign for $\delta_i(x)$. Acceptance of H_0 suggests a unique equilibrium in the data and judgement on the sign of $\delta_i(x)$ is withheld. Using the test statistic $\sqrt{G}(\hat{\mathbf{V}}(\Upsilon_i(x))\hat{\mathbf{\Sigma}}(\Upsilon_i(x))\hat{\mathbf{V}}(\Upsilon_i(x))')^{-1/2}$ $T_{G,i}(\Upsilon_i(x))$, we can choose critical regions at the two tails, each resulting in the rejection of H_0 in favor of either H_+ or H_-.[13] Proofs of consistency and asymptotic levels of the

[13] This is a *directional* hypothesis test. For a recent survey, see Shaffer (2006).

test follow from standard arguments. The player-specific sign tests can also be aggregated according to the procedure in the previous subsection (see, e.g., Shaffer (2006)).

5 Extension: Correlated Private Information

It is possible to relax Assumption 1 regarding (conditional) independence of private information variables across players if one can ascribe groups of observations to the same equilibrium (whenever they have *identical covariates*). We retain the assumption that ϵ_i has positive density over $\mathbb{R}$. We refer to sets of observations playing the same equilibrium as clusters. For example, a market or household observed two or more times or a geographic cluster of games could comprise a cluster. We assume that:

Assumption 3 *Game observations are grouped into (non-singleton) clusters such that a single equilibrium is played within a cluster and chosen (across clusters) according to* Λ_x*.*

In the spirit of Myerson (1991) (pp.371-2)'s remark, a group may be defined by a geographical region or cultural trait. It is also common in the literature to rely on multiple observations of a static game (e.g. market or household) (see Sweeting (2009) or Bajari, Hong, Krainer, and Nekipelov (2010)). As long as equilibria do not change across these observations, they constitute what we call a cluster.[14] Notice nevertheless that different equilibria may be played across clusters. The objective here is to test whether multiple equilibria are indeed played across clusters and use this to infer the sign of interaction effects. For simplicity here, assume that there are only two players and we have access to two observations from a particular cluster with covariate realization equal to x. The idea is to permute the players across games from the same cluster to generate independence of actions when there is only one equilibrium. Observed games are assumed to be iid. Because games within a given cluster follow this equilibrium by assumption, even after the permutation of

[14]Note that, even though these clusters are defined a priori, the single equilibrium assumption within a cluster is also testable by the exact same arguments put forward in Section 3.

players, the same equilibrium will still be played in the (permuted) game. Since the games are independent (even within a cluster), the private signals for two players of different games within a cluster will be independent. Consequently, the permutation allows us to mimic Assumption 1 even if signals are not independent within a game.

Because ϵ's may be dependent within a game, with two players the strategies in equilibrium l are now given by

$$S_i^l(x,\epsilon_i) = \mathbf{1}\left(u_i(x) + \delta_i(x)\mathbb{E}(S_j^l(x,\epsilon_j)|x,\epsilon_i) - \epsilon_i \geq 0\right) \tag{12}$$

for $i \neq j \in \{1,2\}$. Let $\epsilon = (\epsilon_1, \epsilon_2)$ and $\epsilon' = (\epsilon_1', \epsilon_2')$ denote private information variables in different observations from the same cluster. Analogously, the covariates for these two observations are given by X and X'. More concretely, if the cluster consists of observations of a market or household for two periods, primed and non-primed variables will correspond to observations in different periods. Permuting players consists of pairing Firm 1's action in period one to Firm 2's action in period two and vice-versa. Notice that

$$\mathbb{E}(S_1^l(x,\epsilon_1)S_2^l(x,\epsilon_2')|X = X' = x) = \mathbb{E}(S_1^l(x,\epsilon_1)|X = X' = x)\mathbb{E}(S_2^l(x,\epsilon_2')|X = X' = x)$$

since $F_{\epsilon,\epsilon'|X=X'=x}(\cdot,\cdot) = F_{\epsilon|X=x}(\cdot)F_{\epsilon'|X'=x}(\cdot)$. Consequently, when a single equilibrium is played in the data, the (permutation) covariance of actions will be zero (regardless of whether Assumption 1 holds).

If there is more than one equilibrium in the data, the covariance of actions within *permuted* games (i.e. observed actions by pairs of players matched together from different games within a cluster) computed over all observations is

$$\begin{aligned}
\text{cov}(S_1, S_2'|X = X' = x) &= \mathbb{E}\left[\text{cov}(S_1, S_2'|\epsilon_2, \epsilon_1', X = X' = x)|X = X' = x\right] + \\
&\quad +\text{cov}\left(\mathbb{E}[S_1|\epsilon_2, \epsilon_1', X = X' = x], \mathbb{E}[S_2'|\epsilon_2, \epsilon_1', X = X' = x]|X = X' = x\right) \\
&= \mathbb{E}\left[\text{cov}(S_1, S_2'|\epsilon_2, \epsilon_1', X = X' = x)|X = X' = x\right] + \\
&\quad +\text{cov}\left(\mathbb{E}[S_1|\epsilon_2, X = x], \mathbb{E}[S_2'|\epsilon_1', X' = x]|X = X' = x\right) \\
&= \mathbb{E}\left[\text{cov}(S_1, S_2'|\epsilon_2, \epsilon_1', X = X' = x)|X = X' = x\right]
\end{aligned}$$

where $S_i = S_i(x, \epsilon_i)$ and $S'_i = S_i(x, \epsilon'_i)$ and the second and third equalities follow from the fact that, given $X = X = x$, ϵ' and ϵ are independent draws from $F_{\epsilon|X}$. The conditional covariance is

$$\begin{aligned}
&\text{cov}(S_1, S'_2|\epsilon_2, \epsilon'_1, X = X' = x) \\
= &\int \mathbb{E}(S^l_1 \overline{S}^l_2|x, \epsilon_2, \epsilon'_1) d\Lambda_x(S^l) - \int \mathbb{E}(S^l_1|x, \epsilon_2) d\Lambda_x(S^l) \int \mathbb{E}(\overline{S}^l_2|x, \epsilon'_1) d\Lambda_x(S^l) \\
= &\int \mathbb{E}(S^l_1|x, \epsilon_2)\mathbb{E}(\overline{S}^l_2|x, \epsilon'_1) d\Lambda_{\theta,x}(S^l) - \int \mathbb{E}(S^l_1|x, \epsilon_2) d\Lambda_{\theta,x}(S^l) \int \mathbb{E}(\overline{S}^l_2|x, \epsilon'_1) d\Lambda_{\theta,x}(S^l)
\end{aligned}$$

where $S^l_1 = S^l_1(x, \epsilon_1)$ and $\overline{S}^l_2 = S^l_2(x, \epsilon'_2)$. The sign argument now follows if we show that, given $x, \epsilon'_1, \epsilon_2$, if $\delta_i(x) > 0$, $\mathbb{E}(S^l_1(x, \epsilon_1)|x, \epsilon_2) > \mathbb{E}(S^k_1(x, \epsilon_1)|x, \epsilon_2)$ whenever $\mathbb{E}(S^l_2(x, \epsilon'_2)|x, \epsilon'_1) > \mathbb{E}(S^k_2(x, \epsilon'_2)|x, \epsilon'_1)$ (and vice-versa when $\delta_i(x) < 0$). For this, the following assumption is sufficient (though not necessary):

Assumption 4 *For any two equilibria in the support of the equilibrium selection mechanism and any player i, S^l and S^k, either $S^l_i(x, \epsilon_i) \geq S^k_i(x, \epsilon_i)$ for any $\epsilon_i \in \mathbb{R}$ or $S^l_i(x, \epsilon_i) \leq S^k_i(x, \epsilon_i)$ for any $\epsilon_i \in \mathbb{R}$.*

This condition will hold for example under conditional independence of privately observed variables. It also holds when the equilibrium strategies are monotone as assumed in Wan and Xu (2010) and implicitly in Berry and Tamer (2007). A sufficient condition for this is the Single Crossing Property from Athey (2001) (see also Reny (forthcoming)). The assumption can hold more generally though.

In this case, notice that $\mathbb{E}(S^l_2(x, \epsilon'_2)|x, \epsilon'_1) > \mathbb{E}(S^k_2(x, \epsilon'_2)|x, \epsilon'_1)$ implies that $S^l_1(x, \epsilon'_1) \geq S^k_1(x, \epsilon'_1)$ whenever $\delta_i(x) > 0$ (see equation 12). Because of the assumption above, we have $S^l_1(x, \epsilon_1) \geq S^k_1(x, \epsilon_1)$ for every ϵ_1. It should also be that $\{\epsilon_1 : S^l_1(x, \epsilon_1) > S^k_1(x, \epsilon_1)\}$ has positive measure. Suppose this is not the case and $S^l_1(x, \epsilon_1) = S^k_1(x, \epsilon_1)$ for (almost-)every ϵ_1. If this holds, $\mathbb{E}(S^l_1(x, \epsilon_1)|x, \epsilon_2) = \mathbb{E}(S^k_1(x, \epsilon_1)|x, \epsilon_2)$ for any ϵ_2 and consequently $S^l_2(x, \epsilon_2) = S^k_2(x, \epsilon_2)$ implying $\mathbb{E}(S^l_2(x, \epsilon_2)|x, \epsilon_1) = \mathbb{E}(S^k_2(x, \epsilon_2)|x, \epsilon_1)$ and contradicting the original assumption. We then obtain that $\mathbb{E}(S^l_1(x, \epsilon_1)|x, \epsilon_2) > \mathbb{E}(S^k_1(x, \epsilon_1)|x, \epsilon_2)$. Consequently the conditional covariance above is positive. Similarly we can show that the covariance is negative when $\delta_i(x) < 0$. This discussion is summarized in the following proposition:

Proposition 3 *Suppose Assumptions 3 and 4 hold and games are iid. (i) For any given x, multiple BNE exist in the data-generating process if and only if $cov(S_1, S_2'|X = X' = x) \neq 0$; (ii) For all x such that $cov(S_1, S_2'|X = X' = x) \neq 0$,*

$$sign\left(cov(S_1, S_2'|X = X' = x)\right) = sign(\delta_i(x)), \quad i \in \{1, 2\}$$

Since equilibria are allocated to clusters according to the equilibrium selection mechanism Λ_x, this strategy can be implemented using one permuted game from each cluster or a balanced number of permuted games from each cluster. We note nevertheless that even if a different number of permuted games is selected from each cluster, the procedure would still effectively detect multiplicity and identify the signs of the interaction effects, though the mixing distribution in this case will be different from (though dependent on) the equilibrium selection mechanism.

6 Monte Carlo Simulations

In this section we explore Monte Carlo experiments to illustrate the strategy presented in the previous section. The first design reproduces Example 1 and displays multiple equilibria. We use it to analyze the inference procedure on the existence of multiple equilibria and on the interaction signs when more than one equilibrium exists. Design 2 displays only one equilibrium and we use it to illustrate our procedure when multiple equilibria are absent but an excluded variable exists.

Design 1. We study the finite sample performance of the tests for multiple equilibria in Section 4 using a simple design of a 3-by-2 game in Example 1. The design is conditional on some state x and this dependence is suppressed for notational convenience. For some fixed state, let the players' baseline payoffs be $u_1 = 0.5$ and $u_2 = u_3 = 0.3611$, respectively, and let $\delta_i = -1$ and $\epsilon_i \sim \mathcal{N}(\mu = 0.1, \sigma^2 = 0.25^2)$ for all i. Let λ denote the probability with which the first Bayesian Nash equilibrium in (6) shows up in the data-generating process. We experiment with $\lambda = 0.1$, 0.25 or 0.5 and sample sizes $G = 1000$ or 3000.

For any (λ, G), we simulate a data set of players' binary decisions by letting

$$D_{i,g} = 1\left\{u_i - W_g\left(\sum_{j\neq i} p_j^1\right) - (1 - W_g)\left(\sum_{j\neq i} p_j^2\right) - \epsilon_{i,g} \geq 0\right\}$$

where in each game $g \leq G$, W_g is simulated from a Bernoulli distribution with success probability λ, $\epsilon_{i,g}$ from $\mathcal{N}(0.1, 0.25^2)$ and p^ls are propensity-scores in the two Bayesian Nash equilibria. For each (λ, G), we simulate $S = 1000$ data sets. For each data set, we employ the stepwise multiple testing procedure as described in Section 4.2, and make a decision to reject or not to reject the null hypothesis that there is a unique equilibrium in the data-generating process. We experiment with three different approaches for choosing the critical level $\hat{c}_k$ in Section 4.2: (i) simulation using estimated covariance matrix of T_G; (ii) bootstrap; and (iii) studentized bootstrap (Algorithms 3.2 and 4.2 in Romano and Wolf (2005)). For meaningful comparison between these three approaches, we use the same number of simulated multivariate normal vectors in (i) as the number of bootstrap samples drawn in (ii) and (iii) (which is denoted by B). We experiment with $B = 1000, 2000$. In Table 1 below, we report the probability of rejecting at least one true null hypothesis (i.e., rejecting H_0 for $i = 1$) calculated from the $S = 1000$ simulated data sets in columns RP $1, 2, 3$.

Table 2 presents the tests of interaction signs for each of the three players. Since player 1 has the same conditional choice probabilities in the two equilibria, the test withholds judgment for most of the simulations. It detects a negative sign for the other two players.

Design 2. In this design, we consider a 3-by-2-action game where Assumption 2 is satisfied. The baseline payoff for player i is $u_i(x_i) = 1 + x_i$ where $x_1 \in \{-1, 2\}$ and $x_2 \in \{-1/2, 3/2\}$ and $x_3 \in \{-1, 3\}$. Covariate realizations have the same probability. The state-dependent interaction effect for i is $\delta_i(x_i) = \delta x_i$ where δ is a parameter that controls the scale of the interaction effect. The private information ϵ_i is uniformly distributed over $(-c_i, c_i)$, where $c_i = 2(1 + x_i + |\delta x_i|)$.[15] Table 3 lists the marginal choice probabilities, or propensity scores, $p_i(x) \equiv \Pr(i \text{ chooses } 1|x)$ in the unique Bayesian Nash equilibria for each state $x \equiv (x_1, x_2, x_3)$. It is easy to verify that the Bayesian Nash equilibrium is unique for all x from Table 3, since all ϵ_i is uniformly distributed and all propensity scores are strictly between 0 and 1.

To illustrate the non-degeneracy condition of Λ^* in Section 3.2, notice that, when $x_1 = -1$, the equivalence class is $\Upsilon_1(x_1) = \{(-1, -1/2, -1), (-1, -1/2, 3), (-1, 3/2, -1), (-1, 3/2, -1)\}$. In this case, because the equilibrium is unique at each of these four points in Ω_X, $\Lambda_x(\cdot)$ is a degenerate distribution putting probability one on the unique equilibrium for each covariate realization x. Accordingly, $\Lambda^*_{x_i}(p) = 1/4$ if $p \in \{(0.3233, 0.5603, 0.3233), (0.2523, 0.5288, 0.7098), (0.2998, 0.7013, 0.2998), (0.2101, 0.7262, 0.7231)\}$ and is zero otherwise. The important implication is that for each one of these realizations, player 1 adopts a different equilibrium strategy, which implies different a conditional choice probability of choosing 1. As we vary the covariates for the other players while fixing x_1 at -1, we are able to identify the sign of δx_1.

In Design 2, strategic interaction effects are state-dependent and individual-specific. For player 1, states in the first four rows in Table 3 form an equivalence class, while the other four rows form another equivalence class. We simulate $S = 1000$ samples, each with sample size $G = 5000$. For each of these samples, we calculate the test statistics $\hat{\Psi}$ as defined in Section 4 and apply the following decision rule. If $T_{G,i} < -1.64$, then reject H_0 (no interaction effect) in favor of H_- (negative interaction effect). If $T_{G,i} > 1.64$, then reject H_0 in favor of H_+ (positive interaction effect). Otherwise, do not reject H_0. Table 4 below summarizes the finite sample performance of our test. The two entries $[q_+, q_-]$ in the brackets

[15] The parameter c_i is chosen this way to ensure there is a unique Bayesian Nash equilibrium under each state.

report percentages of tests in $S = 1000$ simulations where H_2 is rejected in favor of H_+ (i.e., q_+) and the percentage of rejections in favor of H_- (i.e. q_-), respectively. Recall that the sign of interaction effects for $\delta_i(x_i)$ is the same as the sign of x_i in our design as $\delta > 0$.

7 Empirical Illustration

In this section we investigate how radio stations strategically allocate commercial breaks during their programming schedule. The interaction effects on the payoffs of broadcasting commercials ($\delta(X)$) can be either positive or negative. As explained by Sweeting (2009), if radio stations air commercials at the same time listeners may be dissuaded from switching stations to avoid breaks, and the audience for a particular station is not affected by the decision to broadcast a commercial. On the other hand, if listeners have an outside option (i.e. public radio, a CD, TV), synchronization by all stations risks ultimately driving listeners away, reducing audience for all radio stations. Alternating commercial breaks would in this case be preferable (see Sweeting (2006) for a simple model). Whereas advertisers would like stations to coordinate to preclude consumers from avoiding the ads, radio stations may have an incentive to alternate as ratings are computed on average listenership, not audiences of commercials. Lack of coordination by the radio stations would suggest that the market does not align incentives of advertisers and radio stations.

Sweeting (2009) examined this question by estimating a parametric model. His baseline specifications assumed that (1) stations care symmetrically about their interactions with all other stations in the market and (2) that symmetric equilibria are played. Based on these assumptions, he found that stations prefer to choose the same time for commercials during drivetime hours, with stronger preferences in smaller markets. Our methodology allows us to test whether Sweeting's conclusions are robust to relaxing these possibly restrictive assumptions in a nonparametric setting.

Because programmers have to allocate advertisements in real time (i.e. on the spot) around the usual schedule of songs and news updates without interrupting those pieces of programming, there is uncertainty as to when commercial breaks can be aired. The exact sequence of songs and news updates is not publicly distributed beforehand and, as Sweeting (2009) points out, DJs are given ample discretion over schedules (see footnote 7 in that paper). Therefore we follow Sweeting and assume that the unobserved component of the advertisement timing decision is private information to each radio station.

Warren (2001) mentions that airing commercials at a specific time "can be done some of the time. But it can't be done consistently by very many stations. Few songs are 2:30 minutes long any more" (p.24) (see also Gross (1988)). Hence there is also little reason to believe that this scheduling uncertainty is correlated given public information. This (private) payoff uncertainty to airing a commercial at specific time is captured in our model by ϵ_i.

Given that commercial break choices are made within the one hour programming horizon in real time, whether or not to advertise close to the end of that horizon will not be be affected by continuation value considerations. Furthermore, the number of commercials already aired earlier may induce asymmetries in the payoff to broadcast a commercial at the last minutes of the hour, which are captured by our specification. Data show that most commercials are aired close to the end of the programming horizon (i.e. the hour), so our focus on the end of the hour can also be justified as the relevant empirical focus.

The data sources are BIAfn's *MediaAccess Pro* database, Mediabase 24/7 and the 2001 Census.[16] Based on detailed information on airplay logs for around the first five days of each month in 2001 (59 days in total), the data report the decision of radio programmers to broadcast commercials at minute :55 of four different hours of the day: noon-1pm, 4-5pm, 5-6pm and 9-10pm. We focus on the decision to broadcast commercials at minute :55 or not since this is close to the end of the programming horizon as explained above.[17] Table 5

[16]We thank Andrew Sweeting for providing us the data.

[17]Alternatively, as in some of the specifications used by Sweeting (2009), if the private signal variables follow an extreme value distribution we can restrict our analysis to the choice between :50 and :55 *conditional* on airing commercials at one of these times. Even though we do not impose a particular distributional

depicts the frequency of choices (airing a commercial at :55 or not).

We follow Sweeting (2009) and count music stations as players in the geographic market to which they hold licenses. The specific allocation is done using BIAfn's *MediaAccess Pro* database. There are 6,534 games at the noon-1pm hour, 6,562 games at the 4-5pm hour, 6,536 games at the 5-6pm hour and 6,520 games during the 9-10pm hour. Also available are variables regarding market characteristics. We focus here on the market size obtained from 2001/2 population estimates for individuals aged 12+ reported in BIAfn (based on Census data). For our analysis, we discretize this variable into terciles with the first tercile corresponding to the largest markets.

To best illustrate our methodology, we focus on the three dominant radio stations in each market according to measures of historical listernership. We label players accordingly so that player 1 is the radio station with largest market share, player 2 is the station with the second largest share and player 3 is the station with the third largest portion. The combined market share is on average 41% across all markets, justifying our focus on the strategic interactions among the three largest players. We note that our approach can accommodate a larger numbers of players but we opt for three for illustrative purposes. For example, payoffs can depend on the proportion of competitors choosing to play commercials and not simply on their number (see previous discussion).

Table 6 displays tests of multiplicity conditional on the various hours of the day. We present test results using Wald statistics and the multiple comparison procedure by Romano and Wolf (2005) (with 1,000 bootstrap repetitions) at a targeted 5% FWER. For the RW procedure we show the ordering of the individual test statistics, whether they are positive or negative and which ones are rejected. Unconditionally and conditional on the 4-5pm and 5-6pm hours, we reject the hypothesis of a unique equilibrium and in all three cases this is indicative of a positive strategic interaction effect. We find evidence of multiple

assumption, one can legitimize our procedure as an approximation to a multiple action problem with Extreme Value distributed ϵs. We have also run our procedures using this specification and obtain qualitatively similar results.

equilibrium strategies (across equilibria) for all three players without conditioning on any covariate and for the 4-5pm and 5-6pm hours of the day. Using either procedure (Wald or Romano-Wolf), we are not able to reject the null hypothesis of a unique equilibrium for the hours 12-1pm and 9-10pm. This is in agreement with Sweeting's findings and the fact that larger interaction effects will more likely lead to multiple equilibria. Because listeners are less likely to switch off the radio to an outside option during drivetime hours, radio stations have stronger incentives to coordinate on commercial breaks and retain listenership. In this case, radio stations' incentives are aligned with those of advertisers.

For robustness against the possible failure of the conditional independence assumption, we present permutation versions for the Wald test on Table 7 (see 5). Here we assume that within a given geographical market (i.e. cluster), stations play the same equilibrium in every day of our sample. Different equilibria may nevertheless occur across different clusters. This assumption is also used in Sweeting (2009). Various permutation strategies could have been employed, but for those results player 1 in day d is paired with player 2 in day $d-1$ and player 3 in day $d-2$. There is still evidence of multiplicity for the 4-5pm and 5-6pm hours of the day.

As Sweeting (2009) suggests, smaller markets may present stronger incentives for coordination. Because smaller markets have fewer stations, coordination is easier. Furthermore, if the non-dominant fringe of the market provides more alternatives to listeners as would be the case in larger markets with more stations outside the top-three, the incentives for coordination are not as prevalent. To examine this, we present results conditioning also on terciles of market size. Evidence of multiplicity and positive interaction effects for all players is salient in smallest markets during the 4-5pm and 5-6pm hours of the day but not for the other conditional specifications.

8 Conclusion

In this paper we have shown how a condition typically employed in the analysis of simultaneous games of incomplete information leads to a simple and easily implementable test for the signs of interaction effects and the existence of multiple equilibria in the data-generating process. Inference of the signs of state-dependent and individual-specific interaction effects can be done under minimal assumptions that require only the conditional independence of private information, and the existence of state variables satisfying appropriate exclusion restrictions. Even when the conditional independence of private signals is not in place, we show that identification of signs and detection of multiplicity is possible when the researcher can observe groups of games where players are known to follow strategies prescribed by the same equilibrium. Besides, given that many of the suggested methods for estimating and making inferences in such environments rely on the assumption that only one equilibrium is played in the data, this finding is relevant for the implementation of these techniques.

With discrete covariates, such inference is implementable using well-known results in multiple testing. When a continuous covariate is included, the testing procedure should account for the boundaries between regions with a different number of equilibria. We leave this for future research. Another interesting direction for future research is the inference of interaction effects if strategic dependence exists between games observed in data.

Finally, the conditional independence assumption is also found in dynamic games of incomplete information. In those settings, optimal decision rules involve not only equilibrium beliefs but continuation value functions that may change across equilibria. Though a detailed analysis is deferred to future research, we speculate that our results generalize to such games under certain additional assumptions. In particular, the characterization of optimal policy rules in that context suggests that the existence of a unique equilibrium in the data can still be detected by the lack of correlation in actions across players of a given game as presented in the current paper.[18]

[18]With two actions, the optimal policy for a specific equilibrium would prescribe a decision rule like

References

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

Aradillas-Lopez, A. (2010): "Semi-parametric Estimation of Simultaneous Games with Incomplete Information," *Journal of Econometrics*, 157(2), 409–431.

Arcidiacono, P., and R. Miller (2010): "CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," Working Paper, Duke University.

Athey, S. (2001): "Single Crossing Properties and the Existence of Pure Strategy Equilibria in Games of Incomplete Information," *Econometrica*, 69(4), 861–889.

Bajari, P., J. Hahn, H. Hong, and G. Ridder (forthcoming): "A Note on Semiparametric Estimation of Finite Mixtures of Discrete Choice Models with Application to Game Theoretic Models," *International Economic Review*.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010): "Estimating Static Games of Incomplete Information," *Journal of Business and Economic Studies*, 28(4).

Beresteanu, A., I. Molchanov, and F. Molinari (2009): "Sharp Identification Regions in Models with Convex Predictions: Games, Individual Choice, and Incomplete Data," CeMMAP Working Paper # CWP27/09.

Berry, S., and E. Tamer (2007): "Empirical Models of Oligopoly Entry," in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the*

$S_i(X,\epsilon_i) = \mathbf{1}[u_i(X) + \delta_i(X)\sum_{j\neq i} p_j(X) + \beta\mathbb{E}(V_i(X',\epsilon_i')|X,\epsilon_i,D_i=1) - \epsilon_i \geq \beta\mathbb{E}(V_i(X',\epsilon_i')|X,\epsilon_i,D_i=0)]$ where β is a discount factor, primed variables refer to the following period and $V_i(\cdot)$ is a continuation value defined by a Bellman equation where we make explicit the dependence on equilibrium choice probabilities $(p_1,\ldots,p_N)$ (see for example display (8) in Aguirregabiria and Mira (2007)). If the equilibrium is unique and Assumption 1 holds, conditional choice probabilities will factor as in the static case.

Econometric Society, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge: Cambridge University Press.

Bjorn, P., and Q. Vuong (1984): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation," Caltech Working Paper 537.

Brock, W., and S. Durlauf (2007): "Identification of Binary Choice Models with Social Interactions," *Journal of Econometrics*, 140(1), 52–75.

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

Galichon, A., and M. Henry (2009): "Set Identification in Models with Multiple Equilibria," Working Paper, Université de Montreal.

Grieco, P. (2010): "Discrete Choice Games with Flexible Information Structure: An Application to Local Grocery Markets," Working Paper, Northwestern University.

Gross, M. (1988): "Television: Some Thoughts on Flipping," *Marketing and Media Decisions*, 23(10), 94–6.

Hall, P., A. Neeman, R. Pakyari, and R. Elmore (2005): "Nonparametric Inference in Multivariate Mixtures," *Biometrika*, 92(3).

Honoré, B., and A. de Paula (2010): "Interdependent Durations," *Review of Economic Studies*, 77, 11381163.

Lehmann, E., and J. Romano (2005): *Testing Statistical Hypotheses*. Springer.

Manski, C. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60(3), 531–542.

Moon, R., and B. Perron (2009): "Beyond Panel Unit Root Tests: Using Multiple Testing to Determine the Nonstationarity Properties of Individual Series in a Panel," Working Paper, University of Southern California.

Myerson, R. (1991): *Game Theory: The Analysis of Conflict*. Cambridge: Harvard University Press.

Navarro, S., and Y. Takahashi (2009): "A Semiparametric Test of Agents' Information Sets for Games of Incomplete Information," Working Paper, University of Wisconsin.

Reny, P. (forthcoming): "On the Existence of Monotone Pure Strategy Equilibria in Bayesian Games," *Econometrica*.

Romano, J. P., and A. M. Shaikh (2006): "Stepup Procedures for Control of Generalizations of the Familywise Error Rate," *Annals of Statistics*, 34(4).

Romano, J. P., and M. Wolf (2005): "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, 73(4), 1237–1282.

Seim, K. (2006): "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *The RAND Journal of Economics*, 37(3), 619–640.

Shaffer, J. P. (2006): "Recent Developments Towards Optimality in Multiple Hypothesis Testing," *Lecture Notes-Monograph Series*, 49, 16–32.

Sweeting, A. (2006): "Coordination, Differentiation and the Timing of Radio Commercials," *Journal of Economics and Management Strategy*, 15, 909–942.

——— (2009): "The Strategic Timing of Radio Commercials: An Empirical Analysis Using Multiple Equilibria," *RAND Journal of Economics*, 40(4), 710–742.

Tang, X. (2010): "Estimating Simultaneous Games with Incomplete Information under Median Restrictions," *Economics Letters*, 108(3).

van der Laan, M. J., S. Dudoit, and K. S. Pollard (2004): "Multiple Testing. Part II. Step-Down Procedures for Control of the Family-Wise Error Rate," *Statistical Applications in Genetics and Molecular Biology*, 3(1).

Wan, Y., and H. Xu (2010): "Semiparametric Estimation of Binary Decision Games of Incomplete Information with Correlated Private Signals," Working Paper, Pennsylvania State University.

Warren, S. (2001): *The Programming Operations Manual*. San Marcos, TX: Warren Consulting.

Appendix A: Additional Results

Proof of non-identification of the full structure

Because the data only provide information on the mixtures of equilibria, there are limits to what can be learned about the structure from the data without additional assumptions. This point is illustrated in the appendix using results from the literature on identifiability (or lack thereof) in mixture models.

Let θ denote the structure $(u_i, \delta_i)_{i=1}^N$ and $F_{\epsilon|X}$, and let $\mathcal{L}_{x,\theta}$ denote the choice probabilities profiles corresponding to BNE for a given x and parameter θ. That is, $\mathcal{L}_{x,\theta} \equiv \{p \in [0,1]^N : p \text{ solves (1) for } \theta \text{ and the given } x\}$. We let $\Lambda_{x,\theta}$ be an equilibrium selection mechanism. The following proposition illustrates the limits of what can be learned about the structure from the mixture data without imposing additional assumptions. Let $\#A$ denote the cardinality of set A and define $h : [0,1]^N \longrightarrow [0,1]^N$ as

$$h(p(x); x, \theta) \equiv \left(p_i(x) - F_{\epsilon_i|X}(u_i(x) + \delta_i(x) \sum_{j \neq i} p_j(x)) \right)_{i=1,\ldots,N}. \tag{13}$$

Proposition A1 *Assume*

$$det \left(\frac{\partial h(p(x); x, \theta)}{\partial p(x)} \right) \neq 0$$

Then the structure is not identified if $\#\mathcal{L}_{x,\theta} > \frac{2^N-2}{N}$.

Proof. We first show that, for given x, the number of equilibria is finite. An equilibrium vector $p(x)$ is a fixed point to the mapping depicted on display (1). Equivalently, we represent it as a solution to the following equation:

$$h(p(x); x, \theta) = 0.$$

Notice that $\{0,1\} \cap F_{\epsilon_i|x}(\mathbb{R}) = \emptyset$ for any i, given the full support of ϵ_i. Consequently, for a solution vector, $p_i(x) \notin \{0,1\}$ and $p(x) \in (0,1)^N$. Since

$$\det\left(\frac{\partial h(p(x); x, \theta)}{\partial p(x)}\right) \neq 0$$

the Implicit Function Theorem directly implies that the set of fixed points to (13) is discrete (i.e. its elements are isolated points: each element is contained in a neighborhood with no other solutions to the system). Infinitesimal changes in $p(x)$ will imply a displacement of $h(\cdot; x, \theta)$ from zero, so local perturbations in $p(x)$ cannot be solutions to the system of equations. Since $p(x) \in [0,1]^N$, the set of solutions is a bounded subset of $\mathbb{R}^N$. In $\mathbb{R}^N$, every bounded infinite subset has a limit point (i.e., an element for which every neighborhood contains another element in the set) (Theorem 2.42 in Rudin (1976)). Consequently, a discrete set, having no limit points, cannot be both bounded and infinite. Being bounded and discrete, the set of solutions is finite.

In this case, the observed joint distribution of equilibrium actions is a finite mixture. Given Assumption 1, the cumulative distribution function for the observed actions is given by

$$\Phi(y_1, \ldots, y_N; x, \theta) = \sum_{\mathcal{L}_{x,\theta}} \Lambda_{x,\theta}(p^l(x)) \Pi_{i \in \{1,\ldots,N\}} (1 - p_i^l(x))^{1-y_i}$$

For a given x, the problem of retrieving this cdf and mixing probabilities from observed data is analyzed by Hall, Neeman, Pakyari, and Elmore (2005). In that paper, the authors show that the choice and mixing probabilities ($p_i^l(x)$ and $\Lambda_{x,\theta}$) cannot be obtained from observation of $\Phi(y_1, \ldots, y_N; x, \theta)$ if $\#\mathcal{L}_{x,\theta} > \frac{2^N-2}{N}$. Consequently, it is necessary for identifiability of the relevant probabilities that $\#\mathcal{L}_{x,\theta} \leq \frac{2^N-2}{N}$. Finally, if the equilibrium-specific choice probabilities cannot be identified, the utility function and the distribution of private components cannot be identified either (or else one could obtain the equilibrium specific choice probabilities and use those to obtain the mixing distribution from the data).

■

The condition that $\det\left(\frac{\partial h(p(x);x,\theta)}{\partial p(x)}\right) \neq 0$ is likely to be satisfied. With two players, for example, this determinant equals

$$1 - \delta_1(x)\delta_2(x)f_{\epsilon_1|X}(u_1(x) + \delta_1(x)p_2(x))f_{\epsilon_2|X}(u_2(x) + \delta_2(x)p_1(x)).$$

Also when there are two players, the bound on the number of equilibria implies that, without further assumptions, the existence of more than one equilibrium precludes identification.

Appendix B: A Wald Test for Multiple BNE

By the Delta Method,

$$G^{1/2}(\mathbf{T}_G - \mathbf{\Delta}) \xrightarrow{d} N(\mathbf{0}_N, \mathbf{V\Sigma V'}) \text{ as } G \to \infty$$

where $\mathbf{\Delta} \equiv (\Delta_i)_{i=1}^N$. The Jacobian $\mathbf{V}$ is a N-by-$\tilde{N}$ matrix, with its i-th row V_i defined by the following table (where $\mu_{(m)}, V_{i,(m)}$ denote the m-th coordinates of two $\tilde{N}$-vectors μ and V_i respectively, and $j, k \neq i$),

$\mu_{(m)}$	μ_0	μ_i	μ_j	μ_{ij} or μ_{ji}	μ_{jk}
$V_{i,(m)}$	$\sum_{j\neq i}(-\frac{\mu_{ij}}{\mu_0^2} + \frac{2\mu_i\mu_j}{\mu_0^3})$	$-\sum_{j\neq i}\frac{\mu_j}{\mu_0^2}$	$-\frac{\mu_i}{\mu_0^2}$	$\frac{1}{\mu_0}$	0

Let $\hat{\mathbf{\Sigma}}, \hat{\mathbf{V}}$ be estimates for $\mathbf{\Sigma}, \mathbf{V}$ respectively, constructed by replacing μ_0, μ_I with non-parametric estimates

$$\hat{\mu}_0 = G^{-1}\Sigma_g 1(X_g = x) \; ; \; \hat{\mu}_I = G^{-1}\Sigma_g \left[\Pi_i D_{i,g} 1(X_g = x)\right]$$

Proposition B1 *Suppose the data have G independent games with the same underlying structure and both V and Σ are full-rank. Then*

$$G(\mathbf{T}_G-\mathbf{\Delta})'(\hat{\mathbf{V}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}')^{-1}(\mathbf{T}_G-\mathbf{\Delta}) \xrightarrow{d} \chi^2_{df=N} \quad \text{as } G \to \infty.$$

Under the null, $\mathbf{\Delta} = \mathbf{0}_N$ and the chi-squared distribution can be used to obtain critical values for the test statistic $G\mathbf{T}'_G(\hat{\mathbf{V}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}')^{-1}\mathbf{T}_G$. Because $N \geq 3$ and conditional choice probabilities are bounded away from 0 and 1 (due to the rich support condition in Assumption 1), the full-rank conditions above are not restrictive.

Appendix C: Algorithms for Stepwise Procedure

The following algorithm summarizes the stepwise multiple testing procedure we adopt from Romano and Wolf (2005).

Algorithm C.1 (Basic Non-studentized Step-down Procedure)

Step 1. Relabel the hypotheses in descending order of the test statistics $T_{G,i}$. Let $H^0_{i_k}$ denote the individual null hypothesis whose test statistic is the k-th largest.

Step 2. Set $k = 1$ and $R_1 = 0$.

Step 3. For $R_k + 1 \leq s \leq N$, if $T_{G,(s)} - \hat{c}_k > 0$, then reject the individual null $H^0_{i_s}$.

Step 4. If no (further) null hypotheses are rejected, then stop. Otherwise, let R_{k+1} denote the total number of hypotheses rejected so far (i.e. R_k plus the number of hypotheses rejected in the k-th step), and set $k = k + 1$. Then return to Step 3 above.

We consider two alternatives methodologies for the computation of $\hat{c}_k$: bootstrap and using the asymptotic distribution of the test statistic. The two are summarized in the following two algorithms.

Algorithm C.2 (Computing $\hat{c}_k$ Using Bootstrap)

Step 1. Let i_k and R_k be defined as in Algorithm 1 above.

Step 2. Generate B bootstrap data sets.

Step 3. From each bootstrap data set (indexed by b), compute the vector of test statistics $\left(T^{*,b}_{G,1}, \ldots, T^{*,b}_{G,N}\right)$.

Step 4. For $1 \leq b \leq B$*, compute* $\max^{*,b}_{G,k} = \max_{R_k+1 \leq s \leq N}(T^{*,b}_{G,i_s} - T_{G,i_s})$.

Step 5. Then compute $\hat{c}_k$ *as the* $(1-\alpha)$*-th empirical quantile of the B values* $\{\max^{*,b}_{G,k}\}_{b \leq B}$.

Algorithm C.3 (Computing $\hat{c}_k$ Using Parametric Simulations)

Step 1. Estimate the covariance matrix of the vector of test statistics that corresponds to hypotheses which are not rejected after the first $k-1$ *steps, i.e.* $(T_{G,(R_k+1)}, T_{G,(R_k+2)}, \ldots, T_{G,(N)})$. *Denote the estimate by* $\hat{\Sigma}_k$.

Step 2. Simulate a data set of M observations $\{v_m\}_{m=1}^M$ *from the* $(N-R_k)$*-dimensional multivariate normal distribution with parameters* $(0_{N-R_k}, \hat{\Sigma}_k)$*, where* 0_k *is a k-vector of zeros.*

Step 3. Then $\hat{c}_k$ *is computed as the* $(1-\alpha)$*-th empirical quantile of the maximum coordinates of* v_m *in the simulated data. (M can be large relative to the number of bootstrap samples B in Algorithm A.2 above.)*

The studentized stepwise procedure is summarized in the following algorithm. As before, R_k denotes the total number of hypotheses not rejected in the first $k-1$ steps.

Algorithm C.4 (Studentized Step-down Procedure)

Step 1. Relabel the individual hypotheses in descending order of studentized test statistics $Z_{G,i} \equiv T_{G,i}/\hat{\sigma}_{G,i}$*, where* $\hat{\sigma}_{G,i}$ *are estimates for standard deviation of* $T_{G,i}$.

Step 2. Set $k = 1$ *and* $R_1 = 0$.

Step 3. For $R_k + 1 \leq s \leq S$*, if* $Z_{G,i_s} > \hat{d}_j$*, then reject the individual null* $H^0_{i_s}$.

Step 4. If no further individual null hypotheses are rejected, stop. Otherwise, let R_{k+1} *denote the total number of hypotheses rejected so far and set* $k = k+1$*. Then return to Step 3 above.*

The critical values for the studentized stepwise method $\hat{d}_k$ are computed by an algorithm similar to Algorithm 2.1 where standard errors $\left(\hat{\sigma}^{*,b}_{G,1}, \ldots, \hat{\sigma}^{*,b}_{G,N}\right)$ are also computed in Step 3 and $\max^{*,b}_{G,k} \equiv \max_{R_k+1 \leq s \leq N}(T^{*,b}_{G,i_s} - T_{G,i_s})/\hat{\sigma}^{*,b}_{G,i_s}$ in Step 4.

Table 1: Finite Sample Performance: Tests for Multiple Equilibria

(*Target probability for FWE:* $\alpha = 0.10$)

		$B = 1000$			$B = 2000$		
G	λ	$RP1$	$RP2$	$RP3$	$RP1$	$RP2$	$RP3$
1000	0.50	0.101	0.101	0.095	0.112	0.109	0.111
	0.25	0.093	0.094	0.085	0.094	0.096	0.089
	0.10	0.107	0.107	0.102	0.114	0.119	0.112
3000	0.50	0.108	0.109	0.105	0.087	0.089	0.083
	0.25	0.096	0.097	0.094	0.102	0.105	0.103
	0.10	0.093	0.090	0.092	0.111	0.107	0.108

NOTE: Design 1: Number of simulations $S = 1000$. G is the sample size. λ specifies the probability that the first equilibrium in Example 1 is chosen. *RP1*, *2* and *3* are rejection frequencies of the true null following three tests respectively: (1) the non-studentized test with $\hat{c}_k$ from parametric simulations; (2) the non-studentized test with $\hat{c}_k$ computed via bootstrap; and (3) the studentized test with $\hat{d}_k$ computed via bootstrap.

Table 2: Finite Sample Performance: Test of Signs of Interaction Effects

Brackets include $[q_+, q_-]$.

G	λ	$i = 1$	$i = 2$	$i = 3$
1000	0.50	[0.036, 0.076]	[0.000,1.000]	[0.000,1.000]
	0.25	[0.035, 0.072]	[0.000,1.000]	[0.000,1.000]
	0.10	[0.040, 0.072]	[0.000,1.000]	[0.000,1.000]
3000	0.50	[0.054, 0.067]	[0.000,1.000]	[0.000,1.000]
	0.25	[0.048, 0.048]	[0.000,1.000]	[0.000,1.000]
	0.10	[0.049, 0.053]	[0.000,1.000]	[0.000,1.000]

NOTE: Design 1: S is 1000. G is the sample size. λ is the first equilibrium selection probability. q_+ is the frequency of rejection of H_0 in favor of H_+. q_- is the frequency of rejection of H_0 in favor of H_-.

Table 3: Propensity Scores in Bayesian Nash Equilibria

(p_1, p_2, p_3 in brackets)

x_1	x_2	x_3	$\delta = 0.8$	$\delta = 0.9$	$\delta = 1$
-1	$-1/2$	-1	[0.3233, 0.5603, 0.3233]	[0.3060, 0.5561, 0.3060]	[0.2895, 0.5526, 0.2895]
-1	$-1/2$	3	[0.2523, 0.5288, 0.7098]	[0.2223, 0.5196, 0.7144]	[0.1927, 0.5111, 0.7183]
-1	3/2	-1	[0.2998, 0.7012, 0.2998]	[0.2790, 0.7033, 0.2790]	[0.2590, 0.7048, 0.2590]
-1	3/2	3	[0.2101, 0.7262, 0.7231]	[0.1710, 0.7323, 0.7300]	[0.1316, 0.7376, 0.7360]
2	$-1/2$	-1	[0.7124, 0.5286, 0.2518]	[0.7167, 0.5194, 0.2219]	[0.7203, 0.5109, 0.1922]
2	$-1/2$	3	[0.7479, 0.4754, 0.7477]	[0.7593, 0.4541, 0.7599]	[0.7704, 0.4322, 0.7717]
2	3/2	-1	[0.7249, 0.7263, 0.2098]	[0.7313, 0.7324, 0.1707]	[0.7369, 0.7376, 0.1314]
2	3/2	3	[0.7738, 0.7724, 0.7754]	[0.7927, 0.7903, 0.7955]	[0.8126, 0.8090, 0.8166]

Table 4: Finite Sample Performance: Test of Signs of Interaction Effects

(No. of simulations: $S = 1000$. Brackets include $[q_+, q_-]$.)

	$G = 5000$			$G = 10000$		
	$\delta = 0.8$	$\delta = 0.9$	$\delta = 1.0$	$\delta = 0.8$	$\delta = 0.9$	$\delta = 1.0$
$X_1 = -1$	[0.000, 0.469]	[0.001, 0.628]	[0.000, 0.854]	[0.000, 0.717]	[0.000, 0.890]	[0.000, 0.986]
$X_2 = -1/2$	[0.003, 0.359]	[0.000, 0.520]	[0.000, 0.714]	[0.000, 0.577]	[0.000, 0.790]	[0.000, 0.925]
$X_3 = -1$	[0.000, 0.483]	[0.000, 0.643]	[0.000, 0.834]	[0.000, 0.702]	[0.000, 0.888]	[0.000, 0.986]
$X_1 = 2$	[0.323, 0.004]	[0.459, 0.000]	[0.667, 0.000]	[0.484, 0.000]	[0.736, 0.000]	[0.910, 0.000]
$X_2 = 3/2$	[0.400, 0.000]	[0.617, 0.000]	[0.817, 0.000]	[0.665, 0.000]	[0.867, 0.000]	[0.979, 0.000]
$X_3 = 3$	[0.300, 0.004]	[0.496, 0.000]	[0.735, 0.000]	[0.545, 0.000]	[0.764, 0.000]	[0.930, 0.000]

NOTE: q_+ is the frequency of rejection of H_0 in favor of H_+. q_- is the frequency of rejection of H_0 in favor of H_-.

Table 5: $D_i = \mathbf{1}(:55)$

	Rel. Freq.	Abs. Freq.
1	27.79%	56,653
0	72.21%	21,803

% and number of players choosing :55 or not. D_i is an indicator of whether a commercial is played at :55min.

Table 6: Multiplicity Tests (X=Hour of Day)

		:55min vs not :55min	G
All Hrs.	Wald Test	33.32^*	$26,152$
	RW (2005)	$T_1^\dagger > T_3^\dagger > T_2^\dagger > 0$	
Noon-1pm	Wald Test	3.86	$6,534$
	RW (2005)	$T_3 > T_2 > T_1 > 0$	
4-5pm	Wald Test	13.51^*	$6,562$
	RW (2005)	$T_3^\dagger > T_2^\dagger > T_1^\dagger > 0$	
5-6pm	Wald Test	21.35^*	$6,536$
	RW (2005)	$T_1^\dagger > T_2^\dagger > T_3^\dagger > 0$	
9-10pm	Wald Test	4.23	$6,520$
	RW (2005)	$T_1 > T_3 > T_2 > 0$	

G is the number of games.
$*$: Wald test statistic is significant at 5%.
†: Significant hypothesis at 5% FWER using Romano and Wolf (2005).
T_k is the individual test statistic for player (hypothesis) k. The number of bootstrap repetitions is 1000.

Table 7: Permutation Wald Tests (X=Hour of Day)

	:55min vs not :55min	G
All Hrs.	8.21*	25,850
Noon-1pm	2.20	6,233
4-5pm	27.27*	6,260
5-6pm	10.13*	6,234
9-10pm	1.05	6,218

G is the number of games. Permutations match player 1 in day d to player 2 in day $d-1$ and player 3 in day $d-2$. This accounts for the reduction in sample size.
$*$: Wald test statistic is significant at 5%.

Table 8: :55min vs. not :55min (X=Hour of Day, Market Size)

Market Size (tercile)		Hour of Day 12-1pm	4-5pm	5-6pm	9-10pm
1	Wald Test	0.77	4.94	3.22	2.27
	RW (2005)	$T_3 > T_2 > 0 > T_1$	$T_2 > T_1 > T_3 > 0$	$T_2 > T_1 > T_3 > 0$	$T_1 > T_3 > T_2 > 0$
	G	2,201	2,201	2,200	2,199
2	Wald Test	0.73	3.87	1.97	2.48
	RW (2005)	$T_2 > T_3 > 0 > T_1$	$T_3 > 0 > T_1 > T_2$	$T_2 > T_1 > T_3 > 0$	$T_2 > T_1 > T_3 > 0$
	G	2,157	2,220	2,159	2,153
3	Wald Test	4.96	19.06*	26.07*	2.92
	RW (2005)	$T_2 > T_3 > T_1 > 0$	$T_3^\dagger > T_2^\dagger > T_1^\dagger > 0$	$T_1^\dagger > T_3^\dagger > T_2^\dagger > 0$	$T_1 > T_3 > 0 > T_2$
	G	2,176	2,141	2,177	2,168

$*$: Wald test statistic is significant at 5%.
†: Significant hypothesis at 5% FWER using Romano and Wolf (2005). T_k is the individual test statistic for player (hypothesis) k. The number of bootstrap repetitions is 1000.