INFERENCE FOR GAMES WITH MANY PLAYERS

KONRAD MENZEL†

Abstract. We develop an asymptotic theory for static discrete games ("markets") with a large number of players, and propose a novel approach to inference based on stochastic expansions around a "competitive" limit of the finite-player game. We show that in the limit, players' equilibrium actions in a given market can be represented as a mixture of i.i.d. random variables, where for common specifications identification of structural parameters from the limiting distribution is analogous to static panel models of discrete choice. Our analysis focuses on aggregate games in which payoffs depend on other players' choices only through a finite-dimensional aggregate state at the market level. We establish conditional laws of large numbers and central limit theorems which can be used to establish consistency of point or set estimators and asymptotic validity for inference on structural parameters as the number of players increases. Our results cover games with complete information or private types, as well as intermediate cases, and allow for arbitrary mechanisms for selecting among multiple equilibria.

JEL Classification: C12, C13, C31, C35, C72
Keywords: Multiple Equilibria, Exchangeability, Aggregate Games, Large Games

We consider the problem of estimating, or testing a hypothesis about a parameter that characterizes some aspect of a discrete game. This includes inference regarding structural parameters governing the distribution of payoffs in a population of players, or a descriptive parameter for the reduced form for outcomes of the game. It is generally understood that empirical problems of this type face significant conceptual and technical challenges: for one, strategic interdependence typically makes exact distributions of equilibrium actions difficult to compute and also induces statistical dependence among different agents' choices. Furthermore, economic theory generally does not make unique predictions on players' choices but may admit multiple equilibria. Most existing methods for estimation and inference for game-theoretical models rely on the availability of data for a large number of independent realizations of a game with a given number of players.

Date: February 2012 - this version: April 2014.
† NYU, Department of Economics, Email: konrad.menzel@nyu.edu. This paper was previously circulated under the title "Inference for Large Games with Exchangeable Players." This paper has benefited from comments by seminar audiences at UCL, École Polytechnique, Boston College, Cornell, UCSD, Boston University, Chicago, University of Virginia, Northwestern, Michigan, Stanford, and the North American Summer Meeting of the Econometric Society in Evanston. I would also like to thank Aureo de Paula, Elie Tamer, and Quang Vuong for comments on earlier drafts.

In contrast, this paper develops an asymptotic theory for settings in which the number of observed games ("markets") may be finite or infinite, but the number of players in each game grows large, and is not necessarily the same across markets. We show that under the proposed asymptotic sequence, the limiting distribution of players' choices and characteristics is equivalent to a single-agent discrete choice problem that is augmented by an aggregate equilibrium condition which may still allow for multiple solutions. In analogy to the Cournot oligopoly model with many competing firms, we refer to that asymptotic model as the "competitive limit" of the game. The previous literature considers either the case in which the economy is "in" the limit, or develops econometric techniques based on an exact solution of the finite-player game. In contrast, this paper explicitly derives the first-order approximation error for estimation based on the "competitive" model when the economy is in fact finite and strategic considerations do matter. We also describe a bootstrap bias correction procedure for dealing with the second-order approximation error for games of strategic complements.

We propose inference based on stochastic expansions around the competitive limit that approximate the relevant features of the exact distributions of types and actions for the finite game that is observed in the data. Since that limiting model has a much simpler structure and does not suffer from many of the complications of the finite-player case, we can apply known results to obtain point identification results and derive estimators for the limiting game, and subsequently evaluate the performance of estimators or inference procedures in the finite-player game relative to that limit. Specifically, our results include conditional laws of large numbers (LLN) and a central limit theorem (CLT) with mixing for statistics and estimators that depend on players' (equilibrium) actions. We also illustrate the potential of higher-order bias corrections for estimators to improve their performance in games of moderate size, which can be derived based on the sampling theory we develop in this paper. A simulation study shows that (second-order) bias-corrected point estimators for payoff parameters can perform reasonably well for games with as few as 10 to 15 players. Our results cover games of complete information or private types, as well as intermediate cases. We do not impose any assumptions regarding selection among multiple equilibria in the finite-player economy.

Our analysis focuses on settings in which players' types and actions in the finite game are *exchangeable*. Formally, a sequence of random variables is exchangeable if its joint distribution is invariant to permutations of the ordering of its individual elements. Applied to the context of games, the assumption of exchangeability of players means that at a fundamental level model predictions depend on players' attributes, but not their identities. We will argue that exchangeability of a certain form is a feature of almost any commonly used empirical specification for game-theoretic models with more than two players. In particular, our argument treats only the joint distribution of player types and actions as exchangeable, whereas the resulting conditional distribution of actions given realized types is typically asymmetric. In other words, we only require that players' observed and unobserved characteristics are exchangeable random variables ex ante, but we then allow any aspect of the game - including payoffs, strategies, information structure, and equilibrium selection - to depend on a given realization of players' types in a completely unrestricted manner. Our framework also allows us to treat player-specific parameters as part of a player's type as long as the corresponding parameter space is invariant with respect to permutations of player identities.

The main class of applications analyzed in this paper are aggregate discrete games, in which agents' payoffs depend on the empirical distribution of (or other statistic aggregating) the actions chosen by other players in the market. Models of anonymous strategic interactions of this type are prominent in the empirical and econometric literature on games, including static models of firm entry,[1] models of social interactions and peer effects,[2], or public good provision.[3] We show that we can find an approximate representation of the (complete set of) Bayes Nash equilibria as solutions of a fixed point problem, where both the mapping and the fixed points converge to deterministic limits. A theoretical by-product of our analysis of aggregate discrete games is that, broadly speaking, what information is private or common knowledge among players does not matter for large games to first order. However, we do show that it does affect the asymptotic variance of statistics that depend on players' actions in a market. In that sense information structure affects asymptotic terms up to the order $n^{-1/2}$. Our asymptotic results in section 5 exploit this representation of aggregate games in terms of equilibrium conditions on an aggregate state, an idea which can be applied to other types of games - e.g. Menzel (2013) derives an fixed-point representation of two-sided matching markets with many agents.

Our results cover games of complete information or private types, as well as intermediate cases. We argue that especially when static Bayes Nash equilibrium is used as an empirical model of convenience to characterize agents' choices, but there is some ambiguity regarding the timing of individuals' choices in the true data generating process, a complete information static model may give a more accurate characterization of equilibrium outcomes than assuming private types or any other specific information structure.

Our results are complementary to existing approaches to estimation of games: for one, we find that for the competitive limit structural parameters are typically either point-identified or completely unidentified, whereas for finite games structural parameters are often only

[1]as in Bresnahan and Reiss (1990), Bresnahan and Reiss (1991), Berry (1992), and Ciliberto and Tamer (2009). In Bajari, Hong, and Ryan (2010)'s model for participation in a procurement auction, the strategic aggregate determining expected payoffs in the second-stage auction game is the sum of log c.d.f.s of valuation distributions among the entrants. This aggregate cannot be reduced to a finite-dimensional parameter unless bidder types are discrete and therefore our results do not apply directly in the general case of their model.

[2]see Manski (1993) and Brock and Durlauf (2001)

[3]Todd and Wolpin (2012)

set-identified, where the identified set can be characterized by inequality restrictions on payoffs or choice probabilities.[4] The asymptotic expansions in this paper imply that the diameter of the sharp identification region for the n-player game is typically of the order of n^{-1} or smaller for components that are point-identified in the limit.[5] On the other hand the finite-sample bounds should be expected to become uninformative for larger games for components that are not identified in the limiting game. Hence, our limiting results help understand the identifying power of restrictions for finite games, and assess the width of bounds for structural parameters.

Furthermore, the sampling theory developed in this paper, including a conditional LLN and CLT with mixing, treats players rather than markets as the basic unit of observation. We find that for large games, only information at the level of the individual player (rather than cross-player restrictions) has identifying power regarding structural parameters, so that we can form equality or inequality restrictions that pool player-specific information across different games. Our asymptotic results can then be used to establish asymptotic properties for set estimation or inference based on moment inequalities or equalities for data from a small number of markets, or samples of markets that are heterogeneous in size, assuming that only the number of players in any given market is large.

Related Literature. Our theoretical results apply to games with an intermediate to large number of players, which could be as low as 15 players, depending on the order of approximation. Discrete games of a size at or above that number are a common in empirical studies across several fields. For example, Todd and Wolpin (2012) estimate a model of high school students' and teachers' decisions to exert effort in a classroom of 20 students or more, where students' effort levels are strategic complements in incentivizing teacher effort. While students can choose effort levels on a continuum, there may be multiple equilibria in students' discrete decision whether to provide nonzero effort. In industrial organization, Bajari, Hong, and Ryan (2010) estimate a model of firms' decision whether to participate in a highway procurement auction. The data set used in their analysis records 271 potential entrants that participate in at least one auction.

Large discrete games are also prominent in the literature on social interactions - Soetevent and Kooreman (2007) estimate an interaction model for truancy, smoking, and cell phone use with interaction at the classroom level. Nakajima (2007) considers a model for youth smoking where the reference group consists of other students at the same school. Other

[4]see e.g. Ciliberto and Tamer (2009), Pakes, Porter, Ho, and Ishii (2006), Beresteanu, Molchanov, and Molinari (2011), and Galichon and Henry (2011)

[5]In fact this is a common feature for panel discrete choice models with fixed effects where payoff parameters are known not to be point-identified for a panel of finite length (see Chamberlain (2010)), but consistent estimation is possible as the length of the panel increases and bias correction methods are available for panels of intermediate length, see e.g. Hahn and Newey (2004).

applications of discrete choice models with social interactions in large populations include crime rates where positive interaction effects may arise at the level of a city or neighborhood due to social norms or limited local resources for law enforcement (Glaeser, Sacerdote, and Scheinkman (1996) and Lazzati (2012)), social norms regarding the use of contraceptives in rural communities in a developing country (Munshi and Myaux (2006)), and spillover effects for job search information among unemployed workers (Topa (2001)).

An important aspect of our results concerns the treatment of multiple equilibria. A common approach to estimation of static games of complete information has been either to assume or estimate an equilibrium selection rule,[6] or to construct bounds over the family of all possible mixing distributions.[7] In contrast, for the class of models considered in this paper there is an aggregate state variable for the market that can serve as a finite-dimensional sufficient parameter for equilibrium selection. That market-specific parameter ("fixed effect") can be estimated consistently as n increases, which allows for conditional inference given the selected equilibrium and therefore does not require any assumptions or knowledge regarding the equilibrium selection mechanism. In contrast, unconditional ("random effects") inference would have to account for the equilibrium selection mechanism as a nuisance parameter which in practice requires smoothness or other assumptions on the form of the resulting mixing distribution. Such an approach becomes impractical in large games because both the number of equilibria and the number of covariates potentially affecting equilibrium selection typically increases very fast in the number of players. Furthermore, a nonparametric treatment of the mixing distribution typically requires a large number of observations for each size of the game. Bajari, Hahn, Hong, and Ridder (2011) show that a mixture model of this type can only have a nonzero information bound if the number of possible outcomes of the game is greater than the number of equilibria.

Previous work in the theoretical literature on large games has focussed on purification and approximation properties of distributional games. Kalai (2004) shows that in large populations where types are private information, Bayes perfect Nash equilibrium is ex-post Nash in ε-best responses. For dynamic games, Weintraub, Benkard, and van Roy (2008) show that oblivious equilibrium approximates a Markov-perfect dynamic equilibrium as the number of players grows large. Their work aims at exploiting the computational advantages of working with the large-player limit rather than the finite-player version of the game. In contrast, our focus is on robustness with respect to equilibrium selection, which requires that *any* finite-player equilibrium can be approximated by an appropriately chosen set of equilibria in the limiting game. In this sense, our analysis complements existing theoretical

[6]See e.g. Soetevent and Kooreman (2007), Bajari, Hong, and Ryan (2010), Chen, Tamer, and Torgovitsky (2011), or Todd and Wolpin (2012)

[7]This includes Tamer (2003), Ciliberto and Tamer (2009), Beresteanu, Molchanov, and Molinari (2011), and Galichon and Henry (2011)

results on convergence of finite economies to continuous limiting games. However, it is important to point out that the "competitive limit" for the game typically remains easier to solve, so that our approach is also in part motivated by computational considerations.

Asymptotic approximations (including laws of large numbers and CLTs) that rely on exchangeability rather than the stronger requirement of i.i.d. random variables generally hold only conditional on some appropriately chosen tail sigma-field. Conditional convergence results for infinitely exchangeable arrays - e.g. Blum, Chernoff, Rosenblatt, and Teicher (1958), or Andrews (2005) in econometrics - rely directly on de Finetti's theorem or related ideas. For the results in this paper, we have to construct a coupling for a triangular array of sequences that are only *finitely* exchangeable. Furthermore, we show that for game-theoretic models, the conditioning sigma-field has a structural interpretation in terms of the equilibria of a limiting game, so that uncertainty over a common aggregate state does not preclude valid and robust inference on structural parameters. Our results on aggregative games show some parallels with Shang and Lee (2011)'s conditional analysis of the private information game with many players, and our analysis shows how to adapt their results to the complete information case which does not assume that players actions are i.i.d. conditional on observables.

In the following, we are going to describe the framework we use to model social interactions, and section 3 shows how to establish the conditions for convergence from economic primitives for aggregate games. We then illustrate how to derive moment conditions for estimation and inference from the limiting model in section 4, and section 5 gives generic asymptotic results for sample moments of that type. Section 6 illustrates our main results in a simulation study. Appendix A derives a higher-order bias correction for estimators under many-player asymptotics, and in appendix B we give some further extensions of the baseline specification of our model, allowing for strategic interaction effects that depend on player characteristics in a more general fashion.

Notation: This paper uses standard notation for operations on sets and correspondences. Specifically, we denote the image of a set $A\mathbb{R}^k$ under a mapping $h : \mathbb{R}^k \to \mathbb{R}^l$ with $g(A) := \{y \in \mathbb{R}^l : g(a) = y \text{ for some } a \in A\}$. The Minkowski sum of two sets $A, B \subset \mathbb{R}^k$ is defined as $A \oplus B := \{a + b : a \in A \text{ and } b \in B\}$, and for a vector $x \in \mathbb{R}^k$, we the sum $A + x$ is understood to be the Minkowski sum $A \oplus \{x\}$. Note that if all summands are singletons, Minkowski addition reduces to the standard notion of addition. In order to minimize the notational burden, we adopt the convention from Molchanov (2005) p.195 to use the same standard symbols for addition of sets or (singleton) vectors, where we write $A+B$ for $A \oplus B$, $A - B$ for $A \oplus (-B)$, and $\sum_{i=1}^n A_i$ for the sum $A_1 \oplus A_2 \cdots \oplus A_n$.

A *correspondence* from $\mathcal{X}$ to $\mathcal{Y}$, denoted by $\Phi : \mathcal{X} \rightrightarrows \mathcal{Y}$, is a mapping from elements $x \in \mathcal{X}$ to subsets of $\mathcal{Y}$. We also define the sum of two correspondences $\Phi_1 : \mathcal{X} \rightrightarrows \mathcal{Y}$ and $\Phi_2 : \mathcal{X} \rightrightarrows \mathcal{Y}$

as the Minkowski sum of their images, $\Phi_1(x) \oplus \Phi_2(x)$. The Hausdorff (set) distance between two sets $A, B \in \mathbb{R}^k$ is defined as $d_H(A,B) := \max\{\inf_{a\in A} d(a,B), \inf_{b\in B} d(b,A)\}$ where $d(a,B) := \sup_{b\in B} d(a,b)$ and $d(a,b)$ denotes the Euclidean distance between two vectors $a, b \in \mathbb{R}^k$.

Our notation also distinguishes between vectors $(a_i)_{i\leq n} := (a_1,\dots,a_n)'$ or sequences $(a_i)_{i\geq 1} := a_1, a_2, \dots$ in which elements are ordered, and sets $\{a_i\}_{i\leq n} := \{a_1,\dots,a_n\}$ for which indices do not imply a specific ordering of the elements.

2. Setup

We consider samples that are obtained from M instances of a static game ("market"), where n_m denotes the number of agents in the mth game. Our asymptotic results below require that n_m grows at the same rate across markets, but the number of players may be different across markets at any point in the asymptotic sequence. In order to keep notation simple, we let $n_m = n$ for all markets in the remainder of the paper. Each player i chooses an action $s_{mi} \in \mathcal{S}$ from a set of pure actions $\mathcal{S}$ that is the same for all players and known to the researcher. For the purposes of this paper, we restrict our attention to the case in which agents choose among finitely many discrete actions.

Types and Information: Players' types $t_{mi} = (x'_{mi}, \varepsilon'_{mi})'$ consist of characteristics $x_{mi} \in \mathcal{X}$ that are observed by the econometrician, and unobserved payoff shocks ε_{mi}. We do not restrict the dimension of the random vector ε_{mi}, so that our setup allows for most commonly used finite-dimensional random coefficient models, including alternative-specific taste shifters and heterogeneity in other preference parameters.[8] Each player i observes her own type t_{mi}, and a public signal $w_m \in \mathcal{W}_m$ which is assumed to be common knowledge. We assume that the joint distribution of player-specific information contained in w_m and types $t_{m1}, t_{m2}, \dots$ is infinitely exchangeable, i.e. the conditional distribution of types given the signal satisfies $H_m(t_{m1},\dots,t_{mn}|w_m) = H_m(t_{\pi(1)},\dots,t_{\pi(n)}|w_{m,\pi})$ for any n and permutation π of the set of players $\{1,\dots,n\}$, where we use $w_{m,\pi}$ to denote the transformation of the signal resulting from applying the permutation π to any player-specific information in w_m.

This formulation includes the two polar cases in which w_m is independent of player types (private types), or at the other extreme the case for which $(t_{mi})_{i\geq 1}$ is measurable with respect to the sigma field generate by w_m (complete information). Our setup also permits intermediate cases regarding information structure, for example the signal w_m may contain player specific information about a subvector of t_{mi}. For example the model in Grieco (2012)

[8]Naturally, different identification results will require additional assumptions on the distribution of ε_{mi} - e.g. independence or a parametric model - but the convergence results in this section do not depend on which components of t_{mi} are observed by the researcher.

assumes unobservable taste shifters of the form $\xi_{mi} + \varepsilon_{mi}$, where ε_{mi} is a private signal, and observables x_{mi} and unobserved shocks ξ_{mi} are common knowledge.

For empirical applications, it is important to distinguish between the appropriate informational assumptions for solving the game-theoretic model used for estimation, and what information about types the players are assumed to observe at the outset of the game observed in the data. For example, a simultaneous-move game may be used as a tractable empirical model for a strategic setting in which players may in fact move sequentially or revise previous decisions. In that event, past moves reveal information about players' types, so that any informational assumptions other than complete information may be very sensitive to the exact timing of actions.

For example, complete information static Nash equilibrium can arise from adaptive dynamics in games with strategic complements (Milgrom and Roberts (1990)), or as limit points from strategy revision processes (Blume (1993)) even if players only observe other individuals' past actions, but not their payoffs. On the other hand, giving players the possibility to revise their behavior after learning each others' actions would typically result in movements away from a given realization of a private information static game. Perfect observability of outcomes rather than types is a reasonable assumption in many real-world settings which may exhibit strategic complementarities, e.g. individuals considering whether or not to commit a crime will base their decision on realized crime rates in their community, or school children know how many of their classmates are smokers at a given point in time. Hence if we regard a static game-theoretic model only as an approximation to a stationary point for a process of this kind, Nash equilibria in the complete-information static game would give a more accurate representation of the strategic outcomes than Bayes Nash equilibria for its incomplete-information version.

Actions and Strategies: Each player selects an action from a finite set $\mathcal{S} = (s^{(1)}, \dots, s^{(p)})$, and we let $\Delta\mathcal{S}$ denote the space of distributions over $\mathcal{S}$ represented by the $(p-1)$ dimensional probability simplex endowed with the Euclidean distance $d(\cdot,\cdot)$. Given the information structure of the game, a mixed strategy for player i is given by a measurable map

$$\sigma_{mi} : \begin{cases} \mathcal{T} \times \mathcal{W}_m & \to \quad \Delta\mathcal{S} \\ (t'_{mi}, w'_m)' & \mapsto \quad \sigma_{mi}(t_{mi}, w_m) := \left(\sigma_{mi}(t_{mi}, w_m; s^{(1)}), \dots, \sigma_{mi}(t_{mi}, w_m; s^{(p)})\right)' \end{cases}$$

In the following, we let $\sigma_{-i} = (\sigma_{m1}, \dots, \sigma_{i-1}, \sigma_{i+1}, \dots, \sigma_{mn})$ denote the profile of strategies played by all players except i. We also use the delta function $\delta_{s^{(q)}}$ to denote the distribution that assigns probability 1 to the pure action $s^{(q)}$.

Observable Data: We assume that we observe *type-action characters* $y_{mi} = (s_{mi}, x'_{mi})'$ for a random sample of N_m players (without replacement, possibly the entire market with $N_m = n$) from the markets $m = 1, \dots, M$, where the vector-valued payoff shocks ε_{mi} are not observed by the econometrician but may be known to other players. We denote the *action profile* for the market by $\mathbf{s}_m = (s_{m1}, \ldots, s_{mn}) \in \mathcal{S}^n$, and the *type-action profile* by $\mathbf{y}_m = (y'_{m1}, \ldots, y'_{mn})'$.

Strategic Aggregate: Our formal results focus on aggregate games in which players interact strategically only through a finite-dimensional aggregate state of the economy. For a given strategy profile $\sigma := (\sigma_{m1}, \ldots, \sigma_{mn})$ for the players in the market, we define the aggregate state $G_{mn}(\sigma) := (G_{mn}(s^{(1)}; \sigma), \ldots, G_{mn}(s^{(p)}; \sigma))'$, where

$$G_{mn}(s; \sigma) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\sigma_{mi}(s)|w_m], \qquad s \in \mathcal{S}$$

which is equal to the conditional expectation of the empirical distribution of actions $\hat{G}_{mn} = (\hat{G}_{mn}(s), \ldots, \hat{G}_{mn}(s))'$ where

$$\hat{G}_{mn}(s) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{s_{mi} = s\}$$

$G_{mn}(\sigma)$ and $\hat{G}_{mn}$ coincide in a game of complete information if strategies are pure. In general the strategies σ_{mi} may be mixed, so that the realization of player i's action is not necessarily t_{mi}-measurable.

Appendix B discusses several extensions in which the state variable may be a generalized index of players' types and actions (e.g. entering firms' marginal costs) and include payoff specifications in which interaction effects may be type-specific. Examples for the second case include spatial interactions where the strength of strategic effects depends on the distance between any two players.

Preferences: Interactions among players are assumed to be anonymous in that strategic interdependence is limited to interaction through the aggregate state $G_n(\sigma)$. Specifically, we assume that payoff functions can be written as

$$u_i(s_{mi}, \sigma_{m,-i}, t_{m1}, \ldots, t_{mn}) = u\left(\delta_{s_{mi}}, G_n((s_{mi}, \sigma_{m,-i})), t_{mi}; \theta\right) \tag{2.1}$$

where we let $(\delta_{s_{mi}}, \sigma_{m,-i})$ denote the mixed strategy profile $(\sigma_{m1}, \ldots, \sigma_{i-1}, \delta_{s_{mi}}, \sigma_{i+1}, \ldots, \sigma_{mn})$. In particular, we assume that for any given strategy profile, each agent's payoffs only depend on her own type, but not that of any other players. With a slight abuse of notation, we can then denote player i's expected payoff from choosing action s_{mi} against the profile σ_{-i} by

$$\tilde{u}_i(s_{mi}, \sigma_{-i}|w_m, t_{mi}) := \mathbb{E}\left[u_i(s_{mi}, \mathbf{s}_{-i})|\, w_m, t_{mi}\right] = \sum_{s_{m,-i} \in \mathcal{S}^{n-1}} u_i(s_{mi}, s_{m,-i}) \mathbb{E}\left[\left.\prod_{j \neq i} \sigma_j(t_j, w_m; s_j)\right| w_m\right]$$

Equilibrium: Now, let $\Sigma_{mn} = (\Delta\mathcal{S})^n$ denote the simplex of probability distributions over $\mathcal{S}^n$. The profile $\sigma^*(\cdot, w_m) := (\sigma^*_{m1}(\cdot, w_m), \ldots, \sigma_{mn}(\cdot, w_m))$ is a Bayes Nash equilibrium in the market if with probability 1 we have $\sigma^*_{mi}(t_{mi}, w_m; s) > 0$ only if $\tilde{u}_i(s, \sigma^*_{-i}|w_m, t_{mi}) \geq$

$\tilde{u}_i(s', \sigma^*_{-i}|w_m, t_{mi})$ for all $s' \in \mathcal{S}$. We let

$$\Sigma^*_{mn}(w_m;\theta) := \{\mathbb{E}[\sigma^*(\mathbf{t}, w_m)|w_m] |\, \sigma^*(\cdot, w_m) \text{ is a Bayes Nash equilibrium given} w_m\} \subset \Sigma_{mn}$$

denote the set of distributions of actions that are supported by a Bayes Nash equilibrium given the realized public signal w_m, and players' information (beliefs). Note that $\Sigma^*_{mn}(w_m;\theta)$ depends on the payoff parameter θ through the Nash condition on expected payoffs.

2.1. **Reference Model.** To frame thoughts, consider the following parametric model for a static aggregate game.[9]. Suppose that given the aggregate $G_n(\sigma)$, players have additively separable payoffs of the form

$$u_i(s^{(k)}, \mathbf{s}_{-i};\theta) := \mu_k(x_{mi};\theta) + \sum_{l=1}^{p} \delta_{kl}(x_{mi};\theta) G_n\left(s^{(l)};\sigma\right) + \varepsilon_{mik} \tag{2.2}$$

for $k = 1, \dots, p$. This formulation implicitly assumes that players are risk-neutral with respect to their competitors' actions. Here, the unobserved payoff shifters $\varepsilon_{mi} = (\varepsilon_{mi1}, \dots, \varepsilon_{mip})'$ are independent of x_{mi}, but may be correlated across players. We also assume that the distribution of ε_{mi} and the functions $\mu_k(\cdot)$ and $\delta_{kl}(\cdot)$ are known up to the parameter θ. Players maximize expected utility given their beliefs regarding other players' strategies. The observed action profile $s_1, \dots, s_n$ is then a realization of strategies $\sigma^*_{m1}(t_{m1}, w_m), \dots, \sigma^*_{mn}(t_{mn}, w_m)$ that constitute a Bayes Nash equilibrium together with the beliefs induced by the strategies and the public signal, w_m.

This model is not the most general formulation given our assumptions - e.g. we can allow for more flexible interactions between unobserved characteristics and the other variables. We also do not generally require a parametric model for the systematic parts of payoffs or the distribution of unobservables - section 4.1 discusses estimation and inference on model features relaxing some of these assumptions.

The assumption that strategic interdependence is captured entirely by the aggregate state $G_n = (G_n(s^{(1)}), \dots, G_n(s^{(p)}))$ is common in empirical applications. E.g. Berry (1992) analyzed a static complete information model for entry in airline markets with n potential entrants, where $s_{mi} = 1$ if firm i enters the market, and $s_{mi} = 0$ if it remains inactive. Given a vector x_{mi} of observed firm and market characteristics, his main specification of firm profits is a special case of this setup with $u\left(s, \mathbf{s}_{m,-i}, t_{mi}, (\beta', \Delta)'\right) = s\left(x'_{mi}\beta + \Delta \log\left(nG_n - 1\right) + \varepsilon_{mi}\right)$, where β and Δ are unknown parameters to be estimated from the data. In particular, profits are a function of the number of firms in the market, but not their types or identities.

We can extend this baseline model to allow for strategic effects that are type-specific in that players' decisions may also be influenced by the proportion of agents *of a given type* that choose each action. An extension of this type is discussed in Appendix B. It would also

[9]This model is an adaptation of the nonparametric specification of the private information game in De Paula and Tang (2012)

be straightforward to allow for heterogeneity in individual players' impact on the aggregate - e.g. Ciliberto and Tamer (2009) allow for entry of a larger firm into a market may have a larger effect on other potential entrants' profits than entry of a small competitor.

Manski (1993) and Brock and Durlauf (2001) consider a private information version of a binary choice model of endogenous social effects, in which agents' payoffs depend on the proportion of agents in a reference group choosing either action $s_{mi} \in \{0, 1\}$. They specify payoffs as in the setup of this example with $u\left(s, \delta_{s_{m1}}, \dots, \delta_{s_{mn}}, (x'_{mi}, \varepsilon_{mi})', \theta\right) = s\left(x'_{mi}\beta + \Delta\frac{1}{n}\sum_{j\neq i} s_j + \varepsilon_{mi}\right)$. Common interpretations of this setup in empirical work include models of peer effects, stigma, or contagion.

Equilibrium Selection. In order to describe the set of distributions of type-action profiles $\mathbf{y}$ that are compatible with the solution concept, we introduce an equilibrium selection rule as an auxiliary parameter that characterizes how the ambiguity due to multiple equilibria is resolved. A (potentially stochastic) equilibrium selection rule λ_{mn} is a mapping from w_m to distributions over the equilibrium strategy profiles $\sigma^*_m(\mathbf{t})$. The parameter space for λ_{mn} consists of the mappings of the public signal to the set of mixed Nash equilibria of the game

$$\Lambda_{mn}(\theta) := \left\{\lambda_{mn} : \mathcal{W}_m \to \Sigma_{mn}\,|\lambda_{mn}(w_m;\sigma_m) > 0 \text{ only if } \sigma_m(w_m) \in \Sigma^*_{mn}(w_m;\theta) \text{ a.s.}\right\}. \quad (2.3)$$

where for a given parametrization of random payoff functions, the set of action distributions supported by Nash equilibria, $\Sigma^*_{mn}(w_m;\theta)$, also depends on the parameter θ. In addition, this formulation allows any public information to affect equilibrium selection directly - e.g. if firm size is common knowledge in an entry game, players could always coordinate on an equilibrium in which large firms are more likely to enter than smaller competitors. From the econometrician's perspective, the equilibrium selection rule λ_{mn} in the mth market is a random variable in a cross-section of one or several observed markets.[10]

Given these definitions, the likelihood of the profile for the mth market, $\mathbf{y}_m = (y_{m1}, \dots, y_{mn})$, given $\lambda_{mn} \in \Lambda_{mn}(\theta)$ can be written as

$$f_m(\mathbf{y}_m|\theta, \lambda_{mn}) = \int_{\mathcal{T}^n \times \mathcal{W}_m} \sum_{\sigma^*_m \in \Sigma^*_{mn}(w_m;\theta)} \lambda_{mn}(w_m;\sigma^*_m) \prod_{i=1}^n \sigma^*_{mi}(t_{mi}, w_m; s_{mi}) H_m(dt_m, dw_m),$$

where $H_m(t, w)$ is the joint distribution of t_{mi} and w_m. Hence given $H_m(t, w)$, there is a set of distributions that is indexed with (θ, λ),

$$\mathbf{y}_m \sim f_{y_{m1},\dots,y_{mn}}(y_{m1}, \dots, y_{mn}|\theta, \lambda_{mn}) \quad \theta \in \Theta, \lambda_{mn} \in \Lambda_{mn}(\theta). \quad (2.4)$$

[10]Some of our asymptotic arguments in the next sections require measurability of equilibrium outcomes along a particular filtration, so that the equilibrium selection mechanism has to be defined on a common probability space along sequences of games. However, this does not necessarily restrict the outcome distributions for the game at a given number of players.

We assume that the observed data can be generated by any particular equilibrium (possibly in mixed strategies), where the equilibrium selection rule is not known to the econometrician.

2.2. **Exchangeability.** Asymptotic approximations of the type we just described require that we define sequences of games with an increasing number of players. In order to obtain a fairly general stochastic structure for the game that scales quite naturally, we use the idea of exchangeability: recall that a random sequence $Z_1, \ldots, Z_n$ with joint distribution $f(z_1, \ldots, z_n)$ is said to be *exchangeable* if for any permutation $\pi \in \Pi(1, 2, \ldots, n)$, $f(z_1, \ldots, z_n) = f(z_{\pi(1)}, \ldots, z_{\pi(n)})$. We also say that an infinite random sequence $Z_1, Z_2, \ldots$ is *infinitely exchangeable* if $Z_1, \ldots, Z_n$ are exchangeable for any $n = 1, 2, \ldots$.

Our modeling assumptions are shown to result in (equilibrium) distributions of type-action characters $y_{m1}, \ldots, y_{mn}$ that are exchangeable across agents, where we take the permutations $\pi \in \Pi$ to operate on types, actions, and any player-specific parameters jointly. In particular, the identity of individual agents in a game may often be unknown or irrelevant, so that the parametric family $f_m(\mathbf{y}_m|\theta, \lambda)$, $\theta \in \Theta$ and $\lambda \in \Lambda_{mn}(\theta)$, should be invariant under permutations of the full set of agents, or a known subset, for each instance of the game. If the game has a spatial structure where interactions depend on some notion of distance, the assumption of exchangeability would not be appropriate if the analysis is *conditional* on agents' location. However, unconditional procedures can allow for spatial interaction if agents' locations are endogenous or can be modeled as part of their type t_{mi}, see the discussion of type-specific interactions in appendix B.

While infinitely exchangeable random sequences may be dependent, de Finetti's theorem (see e.g. Theorem 1.1 in Kallenberg (2005)) states that any infinitely exchangeable sequence can be represented as an i.i.d. random sequence from a random marginal distribution.[11] Specifically for any value of n, the joint distribution

$$F_{Z_1,\ldots,Z_n}(z_1, \ldots, z_n) = \int \prod_{i=1}^{n} \hat{F}_{Z_1}(z_i) dQ\left(\hat{F}_{Z_1}\right)$$

where the marginal distribution $\hat{F}_{Z_1}$ distributed according to a measure $Q(\cdot)$ on the space of probability distributions for Z_1. In words, any infinitely exchangeable sequence is a mixture over i.i.d. sequences, where $Z_1, Z_2, \ldots$ are conditionally i.i.d. given the (random) marginal distribution $\hat{F}_{Z_1}$.

Therefore, exchangeability simplifies the characterization of outcomes of social interactions because it allows us to separate cross-sectional heterogeneity from aggregate uncertainty

[11] As an important caveat, note that the equilibrium actions in a game with a finite number of players can in general not be represented as a sample from an infinitely exchangeable sequence, but may only be finitely exchangeable. However, if the empirical distribution of actions for the n-player game converges to a proper limit, Theorem 3.1 in Kallenberg (2005) establishes that the conclusion of de Finetti's theorem remains valid in the limit.

resulting from strategic interdependence or multiplicity of equilibria or random type distributions. Specifically, we have that for n large, the likelihood depends on the equilibrium selection rule λ_0 only through the resulting marginal distribution of type-action characters $y_{mi} = (s_{mi}, x_{mi})$, $f_m^*(y_{m1}|\theta, \lambda) := \frac{1}{n}\sum_{i=1}^n f_m(y_{mi}|\theta, \lambda)$, and we are going to develop a limit theory for relevant features of that distribution as the number of players grows to infinity.

This insight also has implications for identification arguments since it implies that for large markets, any information about the underlying parameters of the population distribution pertains to the (random) marginal distribution $F_\infty(y)$. In general the joint distribution of $\mathbf{y}_m$ contains more information about the parameter θ for any finite-player game than the empirical distribution of individual agents' action-type profiles y_{mi}. Some approaches to identification in finite games rely on features of the joint distribution of the type-action characters across players (e.g. Tamer (2003) or Graham (2008)). However de Finetti's theorem suggests that cross-player features of the joint distribution of the type-action profile may become uninformative about underlying preferences as markets grow "thick," so that in the limit any identifying information on the parameter θ is contained in the marginal distribution of y_{m1}.[12]

Arguments based on exchangeability are also useful for the case in which only a random sample of players (in general without replacement) from a large game is available to the econometrician.[13] Since for large games the joint distribution can be approximated by i.i.d. draws from the empirical distribution of player-level type-action characters, a sufficiently large subsample of players can be used to estimate consistently the empirical distribution arising from the large-scale interaction model, even if it only represents a small share of the agents interacting at the population level.

In order to characterize properties of exchangeable arrays, it is generally useful to define the (descending) filtration $\{\mathcal{F}_n\}_{n\geq 1}$ generated by symmetric events: Given the random sequence $y_{m1}, y_{m2}, \dots$ we say that a random variable Z_n is n-symmetric if it can be written $Z_n = r_n(y_{m1}, \dots, y_{mn}, \dots)$ for some function $r_n(\cdot)$ satisfying $r(y_{m1}, \dots, y_{mn}, \dots) = r(y_{m\pi(1)}, \dots, y_{m\pi(n)}, \dots)$ for any permutation π of $(i)_{i\leq n}$.

Definition 2.1. ***(Tail Sigma-Field)*** *$\mathcal{F}_n$ as the sigma algebra generated by the set of n-symmetric random variables $r(y_{m1}, \dots, y_{mn})$. The tail sigma-field $\mathcal{F}_\infty$ is the sigma algebra*

[12]For example, the main identification argument in Tamer (2003) relies on realizations of types for which subsets of players have a unique dominant strategy, so that the Nash conditions can be characterized as unilateral decisions by the remaining players with several undominated strategies. With a large number of players, there is essentially a "curse of dimensionality" to an identification at infinity argument that relies on simultaneous draws of covariates for a significant fraction of the n players at values for which a given action becomes dominant, so that such a strategy will in general become fragile for large games, and break down in the many-player limit.

[13]E.g. Manski (1993) points out that the relationship between observations in a sample of that type is in general different from that between players in the population from which that sample was drawn.

generated by the set of random variables $r(y_{m1}, y_{m2}, \dots)$ *that are symmetric with respect to any permutation of* $(i)_{i\geq 1}$.

Since any n-symmetric random variable is also $(n-1)$-symmetric, the set of all n-symmetric random variables is contained in the set of all $n-1$-symmetric variables. We therefore have $\mathcal{F}_1 \supseteq \dots \supseteq \mathcal{F}_n \supseteq \mathcal{F}_{n+1} \supseteq \dots \supseteq \mathcal{F}_\infty$. The tail sigma-field $\mathcal{F}_\infty$ also corresponds to the sigma algebra generated by the marginal distributions of y_{m1}. The law of large numbers and central limit theorem below are formulated *conditional* on $\mathcal{F}_\infty$, allowing for the marginal distribution of y_{m1} to remain random even in the limit.

Distribution of Type-Action Characters. One general difficulty with structural inference in large games is that the parameter space Λ_n of equilibrium selection rules - which are functions of all the n players' types - grows very fast in dimension as we let the number of players increase.[14] De Finetti's theorem suggests that, instead of considering the family of resulting joint distributions of the type-action profile in the mth market,

$$f(y_{m1}, \dots, y_{mn}|\theta, \lambda_m) \quad \theta \in \Theta, \lambda_m \in \Lambda_{mn}(\theta),$$

we can restrict our attention to the resulting average marginal distribution of the exchangeable type-action characters, $\frac{1}{n}\sum_{i=1}^n f(y_{mi}|\theta, \lambda_m)$, say. Specifically, if the average marginal distribution of the exchangeable type-action characters converges to a limit $f_m^*(y_{m1}|\theta, \mathcal{F}_\infty)$, the joint distribution of the type action profile becomes

$$\prod_{i=1}^n f_m^*(y_{m1}|\theta, \mathcal{F}_\infty) \quad \theta \in \Theta$$

Note that the dependence of the marginal distribution on the sigma-field $\mathcal{F}_\infty$ indicates that in general the limiting marginal distribution $f_m^*(y_{m1}|\theta, \mathcal{F}_\infty)$ is a random object whose distribution no longer depends on market size, n. There are two main reasons why the limit may remain non-deterministic: for one, types $t_{m1}, t_{m2}, \dots$ were assumed to be exchangeable but not necessarily i.i.d., so that the marginal type distribution in the mth market $H_m(x, \varepsilon)$ is in general not fixed. On the other hand, the equilibrium selection rule λ_m governing the data generating process may be random and may randomize between distinct equilibrium distributions even in the limit. At this point, we leave the conditioning set $\mathcal{F}_\infty$ unspecified, but we will see in the next section that for aggregate games it can be conveniently represented by the sigma field generated by the type distribution $H_m(x, \varepsilon)$ and limiting equilibrium value of the aggregate which may only take a finite number of different values.

[14]The generic maximal number of pure Nash equilibria in an n-player discrete game with p strategies is p^{n-1} (see McLennan (1997)).

3. Convergence of Equilibria

Since the main goal of this paper is to ensure that the proposed methods for inference are asymptotically valid and robust to equilibrium selection, we have to ensure that (1) the data generating process always has a well-defined limit, and that (2) the asymptotic approximation does not eliminate or ignore any of the equilibria in the finite-player economy. In this section we derive limiting results for aggregate games. For ease of exposition, we develop our results for a given market m, and omit the first subscript m in the following discussion whenever this does not lead to any ambiguities.

3.1. **Fixed-Point Representation.** Given payoff functions, we define the set of best responses to an aggregate profile G as

$$\psi_0(t;G) := \text{conv}\left\{\delta_s, s \in \mathcal{S}\,|u(s,G,t) \geq u(s',G,t) \text{ for all } s' \in \mathcal{S}\right\}$$

where δ_s is the unit vector in $\mathbb{R}^p$ corresponding to action $s \in \mathcal{S}$. In words, $\psi_0(t;G)$ represents the probability distributions over the pure strategies that are best responses to a value of the aggregate equal to G for a player of type t.

At every value of G, $\psi_0(t_{mi};G)$ is singleton with probability 1 in most standard applications, in which case the expected best-response

$$\Psi_{m0}(G) := \mathbb{E}\left[\psi_0(t_{mi};G)\right] \tag{3.1}$$

is a (single-valued) function of G. Assumption 3.3 below gives primitive conditions on payoff functions and the distribution of t_{mi} for $\Psi_{m0}(G)$ to be a continuous function of G. The assumption that $\Psi_{m0}(G)$ is a function rather than a set-valued mapping is not needed for our main convergence results but simplifies the exposition substantially. For a full discussion of the more general case, we refer the reader to Appendix B of this paper.

For the n-player game, each player has to account for the effect of her strategy on the aggregate G_n. To this end it is useful to consider the average behavior among all players except i. Specifically, consider player i's choice of $\sigma_{mi}(t,w) = \sigma$ for a fixed $t \in \mathcal{T}$ and suppose that her Bayes strategy for values $t' \neq t$ is given by $\sigma^*_{mi}(t', w_m) \in \psi_0(t';G)$. The resulting value of the aggregate is then given by

$$G_{mn} = \frac{1}{n}\sum_{j\neq i}\mathbb{E}[\sigma_{mj}(t_{mj},w_m)|w_m] + \frac{1}{n}\left\{\mathbb{E}[\sigma^*_{mi}(t_{mi},w_m)\mathbb{1}\{t_{mi}\neq t\}|w_m] + \sigma P(t_{mi}=t|w_m)\right\}.$$

We therefore define

$$\tilde{G}_{-i,n}(G,t,w_m;\sigma) := \frac{n}{n-1}G - \frac{1}{n-1}\left\{\mathbb{E}[\psi_0(t_{mi};G)\mathbb{1}\{t_{mi}\neq t\}|w_m] + \sigma P(t_{mi}=t|w_m)\right\}$$

for $t_{mi} \in \mathcal{T}$ and $\sigma, G \in \Delta\mathcal{S}$. If the conditional distribution of $t_{mi}|w_m$ has a continuous component and satisfies certain other standard regularity conditions, then Lemma 3.2 below

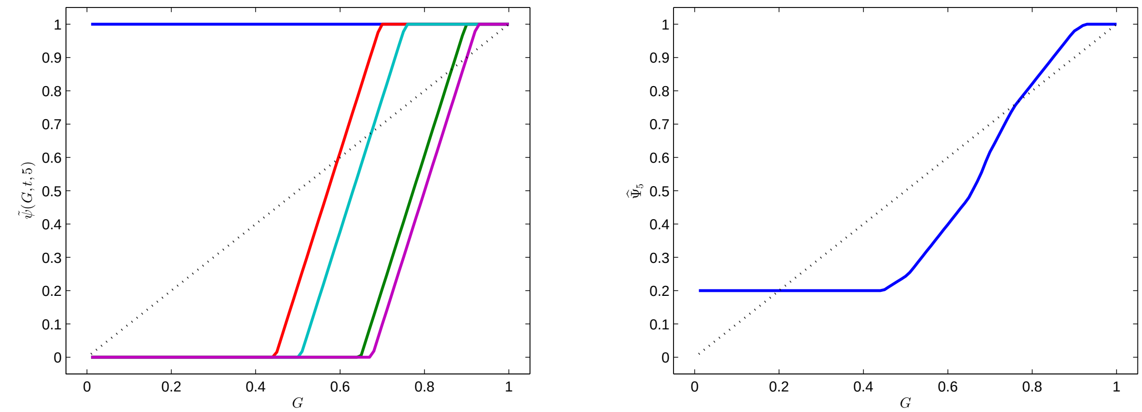

FIGURE 1. The correspondence $\psi_5(t_{mi}; G)$ for five draws of t_{mi} in a five-player binary action game (left), and the corresponding aggregate response mapping $\hat{\Psi}_5(G)$ (right).

implies that the set expectation in the second term is a singleton, and furthermore $P(t_{mi} = t|w_m) = 0$. However, this formulation also includes the case of a complete information game in which $P(t_{mi} = t|w_m) = 1$.

In order to characterize the equilibrium distributions of actions $G_n^* = (G_{mn}^*(s^{(1)}), \dots, G_{mn}^*(s^{(p)}))'$, we define the correspondence of implicit best responses supporting an aggregate state G as

$$\psi_n(t; G) := \left\{\sigma \in \Delta\mathcal{S} \,|\, \sigma \in \psi_0(\tilde{G}_{-i,n}(G, t, w_m; \sigma), t) \text{ and } \tilde{G}_{-i,n}(G, t, \sigma) \in \Delta\mathcal{S}\right\}$$

Also, let the aggregate response mapping

$$\hat{\Psi}_{mn}(G) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\psi_n(t_{mi}; G)\right| w_m]$$

Figure 3.1 illustrates the construction of $\hat{\Psi}_{mn}(G)$ from the individual best responses $\psi_n(t_{mi}; G)$ for a complete information binary action game, given a realization of types $t_{m1}, t_{m2}, \dots$.

The following Proposition characterizes the set of (Bayes) Nash equilibria of the n-player game as fixed points of $\hat{\Psi}_{mn}$:

Proposition 3.1. *Suppose payoff functions are of the form (B.1). Then there exists a Bayes Nash equilibrium with distribution G_{mn}^* if and only if $G_{mn}^* \in \hat{\Psi}_{mn}(G_{mn}^*)$.*

See the appendix for a proof.

3.2. **Assumptions.** The main convergence result in this section establishes that the values of G_{mn}^* that are supported by Nash equilibria in the finite game, $G_{mn}^* \in \hat{\Psi}_{mn}(G_{mn}^*)$, converge to fixed points of the expected best response mapping, $G_{m0}^* \in \Psi_{m0}(G_{m0}^*)$. Our argument is based on convergence of the equilibrium mappings, $\hat{\Psi}_{mn}(G)$ to $\Psi_{m0}(G)$ and requires two technical conditions: for one, Assumption 3.1 below gives a primitive condition on payoffs that is sufficient for stochastic convergence to be uniform in G. In addition, we assume that the empirically relevant fixed points of the limiting mapping $\Psi_{m0}(G)$ are regular in the sense specified in Assumption 3.2 below in order for uniform convergence of the mappings to imply convergence of their fixed points.

In order to characterize the stochastic properties of $\psi_0(t_{mi}; G)$, define the set of types for which s is a best response to G by

$$A(s, G) := \{t \in \mathcal{T} : u(s, G, t) \geq u(s', G, t) \text{ for all } s' \in \mathcal{S}\}$$

Our main assumption on the distribution of payoffs restricts the degree to which $A(s, G)$ may vary as we change the value of the aggregate state G:

Assumption 3.1. *(i) The payoff functions are of the form (B.1), and (ii) the collection*

$$\mathcal{A} := \{A(s, G) : s \in \mathcal{S}, G \in \Delta\mathcal{S}\}$$

is a VC class of sets.

There are a number of empirically relevant examples for which the second part of Assumption 3.1 is satisfied: Clearly, $\mathcal{A}$ is a VC class if the number of types t is finite as in Kalai (2004). Also if payoffs have a linear index structure in t_{mi} for all G, i.e.

$$u(s, G, t) = v(s, t'\beta(G), G)$$

then the sets $A(s, G)$ are intersections of the type space $\mathcal{T}$ with linear half-spaces in $\mathbb{R}^{\dim(t)}$, and therefore a VC class with index less than or equal to $\dim(t) + 1$. This example is a generalization of the framework considered by Brock and Durlauf (2001).

We next state a regularity condition that ensures that the relevant solutions to the limiting equilibrium mapping Ψ_{m0} defined in (B.2) are stable under perturbations of Ψ_{m0}, so that the set of equilibria remains stable along sequences of mappings approximating the limiting mapping. This condition also implies that solutions to the fixed point problem $G \in \Psi_{m0}(G)$ are locally unique, from which it can be shown that the number of distinct equilibrium distributions is finite.

Assumption 3.2. ***(Regular Economy):*** *For every $\theta \in \Theta$ and population distribution $H_0 \in \mathcal{P}$, $\Psi_{m0}(G)$ is single-valued for all $G \in \Delta\mathcal{S}$ and one of the following holds:*

(i) *At every distribution G^*_{m0} solving $G^*_{m0} \in \Psi_{m0}(G^*_{m0})$, the Jacobian $\nabla_G \Psi_{m0}(G^*)$ is defined, and $I_p - \nabla_G \Psi_{m0}(G^*)$ has rank equal to $p - 1$, or*

(ii) *If G^*_{m0} is a cumulation point of the sequence $(G^*_{mn})_{n \geq 1}$, then with probability 1, $\nabla_G \Psi_{m0}(G^*_{m0})$ is defined, and $I_p - \nabla_G \Psi_{m0}(G^*_{m0})$ has rank equal to $p - 1$*

where I_p denotes the p-variate identity matrix.

If a fixed point $G^*_{m0} = \Psi_{m0}(G^*_{m0})$ satisfies the maximal rank condition in part (i), we also say that G^*_{m0} is *regular*, and we define

$$\mathcal{G}^*_{m0} := \{G^*_{m0} : G^*_{m0} = \Psi_{m0}(G^*_{m0}) \text{ and } G^*_{m0} \text{ is a regular point of } \Psi_{m0}(G)\}$$

as the set of regular fixed points of Ψ_{m0}. Note that the alternative conditions (i) and (ii) are nested, and our formal results only require the weaker version, part (ii), to hold. Rank conditions for existence and local stability of equilibria are standard in econometric analysis of equilibrium models,[15] and violations of a generalized rank condition of this kind are typically "non-generic" in the sense that they correspond to subsets of the parameter space of (Lebesgue) measure zero. In this paper we do not derive primitive conditions for this assumption in terms of the population distribution of types and the parameter space Θ.

Non-regular fixed points of Ψ_{m0} are generally unstable both with respect to local perturbations of the mapping Ψ_{m0} as well as best-response dynamics under local perturbations of the corresponding equilibrium. In the simulation experiments for section 6, we do in fact find that for small markets, properties of higher-order bias corrections deteriorate substantially for specifications that are close to violating this assumption. These considerations are particularly important if the type distribution $H_m(t)$ is random at the market level - as in the case of correlated types - which is the scenario for which we rely on the requirement in part (ii).

Of the alternative assumptions on genericity of equilibrium points, only part (iii) restricts equilibrium selection in the finite-player game by assuming that only regular fixed points of Ψ_{m0} are empirically relevant as limits for the equilibria in the observed markets. While there is no generally accepted theory for equilibrium selection, this requirement can be motivated through local stability properties of equilibrium points: Note that if the Hessian of Ψ_{m0} is not degenerate at G_0, then non-regular fixed points of the best-response mapping correspond to equilibria that are locally unstable with respect to best-response dynamics under local perturbations in certain directions. If non-regular points can be ruled out, as in parts (i) or (ii), our results to not require any restrictions on equilibrium selection.

One useful implication of the assumption of a regular equilibrium mapping is that the number of fixed points of Ψ_{m0} is finite:

Lemma 3.1. *Under Assumption 3.2 part (i), the number of fixed points solving $G^*_{m0} \in \Psi_{m0}(G^*_{m0})$ is finite. Furthermore, under either alternative of Assumption 3.2 (a), the cardinality of $\mathcal{G}^*_{m0}$ is finite.*

See the appendix for a proof, which is very similar to the classical arguments for finiteness of the number of equilibria in regular market economies (e.g. Proposition 17.D.1 in Mas-Colell, Whinston, and Green (1995)) or finite-player games (Theorem 1 in Harsanyi (1973)).

[15]See e.g. Assumption 2.1 in Hausman (1983) for the linear simultaneous equations model.

While Assumptions 3.1 and 3.2 are sufficient for convergence of the equilibrium values of G^*_{mn}, the law of large numbers and central limit theorem in the next section require that we define the equilibrium actions for each game along the asymptotic sequence on a common probability space. Our argument establishing such a coupling requires that the joint distribution of best responses varies continuously in the aggregate G and that the expected best response $\Psi_{m0}(G)$ is a single-valued function. More specifically, we impose the following set of additional assumptions on preferences and types for the results requiring a minimal degree of smoothness:

Assumption 3.3. *Types can be partitioned as $t_{mi} = (t'_{mi1}, t'_{mi2})'$ such that (i) the conditional distribution of t_{mi1} given t_{mi2} is continuous with a uniformly bounded continuous p.d.f., and (ii) either of the following holds:*

(a) ***("Unordered Choice")*** *the subvector t_{mi1} has full support on $\mathbb{R}^{p-1}$, and for the payoff vector $\mathbf{u}(G,t) := \left(u(s^{(1)}, G, t), \ldots, u(s^{(p)}, G, t)\right)$, we have that $\nabla_{t_{m1}}\mathbf{u}(G,t)$ is continuous in t and has rank $p-1$ for all $t \in \mathcal{T}$ and $G \in \Delta\mathcal{S}$.*

(b) ***("Ordered Choice")*** *$t_{mi1} \in \mathbb{R}$, and $u(s^{(l)}, G, t) - u(s^{(k)}, G, t)$ is strictly monotone in t_{m1} for $k \neq l$ and almost all $t \in \mathcal{T}$.*

(iii) x_{mi} is a subvector of t_{mi2}.

If we consider payoffs for a fixed value of G, either condition (a) or (b) are met by most standard parametric random utility models for choice among discrete alternatives. Typically, specifications include additive alternative-specific taste shifters that are continuously distributed and have unbounded support conditional on observable characteristics. While part (b) of the assumption does not require a particular ordering of the alternative choices, classical specifications for ordered choice models include a scalar source of heterogeneity that satisfies the monotonicity condition on utility differences. From a technical standpoint, Assumption 3.3 also greatly simplifies the analysis as the following lemma shows.

Lemma 3.2. *Suppose Assumption 3.3 (i) and (ii) hold. Then for almost all types t_{mi}, $\psi_0(t_{mi}; G)$ is single-valued, and the expected best-response $\Psi_{m0}(G)$ is a single-valued continuous function at all values of $G \in \Delta\mathcal{S}$.*

The proof for this lemma is in the appendix. This result helps avoid tedious case distinctions in the characterization of sequences that approach mixed strategy equilibria in the limiting game when we turn to the construction of a coupling between games with different numbers of players below.

3.3. **Convergence of Equilibria.** Our main convergence result in this section establishes that the set of fixed points

$$\mathcal{G}^*_n := \left\{G^* : G^* \in \hat{\Psi}_{mn}(G^*)\right\}$$

approaches the set of regular solutions to $G = \Psi_{m0}(G)$, which implies that the set of (Bayes) Nash equilibria in the finite game can be approximated by a subset of $\mathcal{G}^*_{m0}$. We can now state our main result on stochastic convergence of the equilibrium values of the aggregate G^*_{mn}:

Theorem 3.1. *Suppose Assumptions 3.1 and 3.2 hold, and that G^*_{mn} is a sequence of empirical distributions solving $G^*_{mn} \in \hat{\Psi}_{mn}(G^*_{mn})$. Then we have that (a) $d(G^*_{mn}, \mathcal{G}^*_{m0}) \stackrel{a.s.}{\to} 0$, and (b) with probability approaching 1, for every $G^*_{m0} \in \mathcal{G}^*_{m0}$ and every neighborhood $B(G^*_{m0})$ of G^*_{m0} we can find $\tilde{G}_n \in B(G^*_{m0})$ such that $\tilde{G}_n \in \hat{\Psi}_{mn}(\tilde{G}_n)$.*

See the appendix for a proof. Given the conclusion of 3.1, part (a) of the conclusion implies that although the number of distinct equilibria may grow very fast as the number of players increase, the implied distributions of actions become concentrated near a finite set $\mathcal{G}^*_{m0}$. Taken together, parts (a) and (b) show that the set $\mathcal{G}^*_{m0}$ of equilibrium values for the population correspondence Φ_0 is equal to the set of limiting points of the set of equilibria in the n-player games. Note that the full-rank condition in Assumption 3.2 is crucial in attaining the stronger conclusion (b).

It is straightforward to check that the solutions of the fixed-point problem $G \in \Psi_{m0}(G)$ correspond to the Bayes Nash equilibria (BNE) for a version of the game with $n \geq 2$ players in which types t_{mi} are private knowledge. Since by definition, individual actions arising from a BNE only depend on a player's own type, from the individual agent's perspective, the informational requirements for optimal play are much lower in the limiting game than in the complete information version with finitely many players. Since the mapping Ψ_{m0}, and therefore its fixed points $\mathcal{G}^*_{m0}$ do not depend on the information structure of the finite-player game, our results imply that those differences are less important in aggregate games with a large number of agents.

3.4. **Coupling.** Our main asymptotic results concern almost sure convergence of $(y_{mi,n})_{i \leq n}$ to a limiting sequence of type-action profiles $(y_{mi})_{i \geq 1}$, where $G^*_{m0} = \mathbb{E}[y_{m1}|\mathcal{F}_\infty]$ and

$$y_{mi} \in \psi_0(t_{mi}; G^*_{m0}) \qquad \text{a.s.}$$

where under Assumption 3.3, $\psi_0(t_{mi}; G^*_{m0})$ is single-valued with probability one. In that case, $(y_{mi})_{i \geq 1} = (\psi_0(t_{mi}; G^*_{m0}))_{i \geq 1}$ also constitute an infinitely exchangeable array since types t_{mi} are also infinitely exchangeable. Since the set of limiting equilibria $\mathcal{G}^*_{m0}$ may have more than one element, convergence of the set of possible equilibria alone does not guarantee convergence for a particular sequence of a given player i's actions. Instead, for any realization of the payoffs the economy could "cycle" between distant equilibrium values for G^*_{mn} as n increases.

Denoting player i's type-action character in the n-player game with $y_{mi,n} := (s_{mi,n}, x_{mi})$, a meaningful notion of almost sure, or conditional convergence therefore requires that the triangular array $(y_{m1,n}, \dots, y_{mn,n})_{n\geq 1}$ and the limiting sequence $(y_{mi})_{i\geq 1}$ are defined on a common probability space. To address that difficulty, we introduce a coupling of the sequence of type-action characters determined in equilibrium, $y_{m1,n}, \dots, y_{mn,n}$, to the infinitely exchangeable sequence $y_{m1}, y_{m2}, \dots$.

Definition 3.1. ***(Coupling)*** *Suppose that for a fixed value n_0, the type action character y_{mi,n_0} is distributed with marginal p.d.f. $f_{m,n_0}(y|\theta)$. We then say that $(y_{mi,n_0})_{i\leq n_0}$* can be embedded into a triangular array of equilibrium outcomes $(y_{mi,n})_{i\leq n}$ *converging to an infinitely exchangeable array $(y_{mi})_{i\geq 1}$ if there exists a sequence of equilibrium selection rules $\lambda \in \Lambda_n(\theta)$ such that (i) $f_{m,n_0}(s, x|\theta, \lambda_{n_0}) = f^*_{m,n_0}(s, x|\theta)$ for all $s \in \mathcal{S}^n$, (ii) the type action characters $y_{m1,n}, \dots, y_{mn,n}$ are exchangeable for each n, and (iii) there exists a deterministic null sequence $c_n \to 0$ such that for any bounded function $m(y, G; \theta)$,*

$$|\mathbb{E}\left[m(y_{m1,n}, G; \theta) - m(y_{m1}, G; \theta)\, |\mathcal{F}_n\right]| < c_n \quad \text{a.s.}$$

for all n, $\theta \in \Theta$, and $G \in \Delta\mathcal{S}$.

This definition of a coupling of equilibrium outcomes of the game stipulates that the sequence match the cross-sectional distribution of type-action characters at market size n_0. We show below that when $f_{m,n_0}(y|\theta)$ is generated by a mixture over Bayes Nash equilibria, such a coupling can be constructed with a bounding sequence c_n in part (iii) that is independent of the distribution of equilibria in the n_0 player game.

It is important to distinguish the role of the coupling in our analysis from a factual description of "real-world" economic behavior if more players were added to an existing market. For the asymptotic results below, an asymptotic sequence is constructed for the sole purpose of bounding a statistical approximation error, so that a uniform error bound of this type is sufficient to ensure robustness with respect to equilibrium selection. Since the game is only observed at one fixed value of $n = n_0$, the coupling of the random elements $y_{mi,n}$ across different values of $n > n_0$ has no empirical significance.

For aggregate games, we show that such a coupling can constructed by introducing an auxiliary unobserved type $\nu_{m1}, \nu_{m2}, \dots$ that governs equilibrium selection and is independent of the payoff-relevant types $t_{m1}, t_{m2}, \dots$. Specifically, we let $(\nu_{mi})_{i\geq 1}$ be an infinitely exchangeable sequence of random variables in $\mathbb{R}^p$ that are independent of $(t_{mi})_{i\geq 1}$, so that the augmented types $(x_{mi}, \varepsilon_{mi}, \nu_{mi})_{i\geq 1}$ also form an infinite exchangeable sequence. We assume without loss of generality that the states determining equilibrium selection are public information, i.e. the states $\nu_{m1}, \nu_{m2}, \dots$ are w_m-measurable. In economic terms, this implies that in all states of the world, players agree on which equilibrium is being played.

Given the augmented type, we assume that the sequence of (random) equilibrium selection rules $\{\lambda_n\}_{n\geq 1}$ is of the form $\lambda_{ni}(w_m) := \tilde{\lambda}_n(\nu_{mi}, w_m)$ and invariant to permutations of the agent-specific information in w_m, where $\lambda_n \in \Lambda_n(\theta)$ the set of equilibrium selection rules defined in (2.3). This formulation ensures that the resulting distribution of type-action characters $y_{mi,n}$ is exchangeable for every n. However for any given realization of $\nu_{m1}, \dots, \nu_{mn}$, λ_n need not be symmetric, and may also select asymmetric equilibria.

The following proposition shows that imposing this structure on λ_n is not restrictive in terms of the joint distributions of type-action profiles that can be generated for fixed n.

Proposition 3.2. ***(Coupling for Aggregate Games)*** *Suppose Assumptions 3.1-3.3 hold. For the n_0-player game assume that the observed action profile is generated by a mixture over the Nash equilibria for the type-profile $(t_{mi})_{i\leq n_0}$ with probability one, and let $f^*_{m,n_0}(s,x|\theta)$ be the resulting unconditional distribution over type-action characters. Then $(y_{mi,n_0})_{i\leq n_0}$ can be embedded in a triangular array of equilibrium outcomes converging to (y_{mi}) $i\geq 1$, where the bounding sequence c_n does not depend on $(t_{mi})_{i\leq n_0}$ or the distribution of equilibria generating $f^*_{m,n_0}(s,x|\theta)$.*

The proof for this result is in the appendix. Note that the conclusion of the proposition 3.2 implies that a coupling of this type can be found for any possible cross-sectional distribution over Nash equilibria in the n-player game.

4. Estimation and Inference

This section develops strategies for estimation and inference that combine data from the finite player game with moment restrictions derived for the limiting model. In general, there is a qualitative difference between identification of structural parameters from the exact finite-player distributions for a given value of n and identification from the "competitive limit" under this asymptotic sequence. Since this paper considers consistency of estimators and validity of inference as the number of players increases, we need to consider identification from the limiting distribution $f(y_{mi}|\mathcal{F}_\infty)$ for large-sample results rather than their analogues for the finite-player problem.[16]

4.1. **Moment Conditions.** We formulate our main asymptotic results in terms of moments of the type-action characters, $m(y; G, \theta)$, where the population moment is given by

$$m_0(\theta) := \mathbb{E}[m(y_{mi}, G^*_{m0}; \theta)|\mathcal{F}_\infty]$$

[16]For example, the identification condition in a generic consistency result for extremum estimators typically requires that there exists a "population" criterion function $Q_0(\theta)$ that is uniquely minimized at the true parameter, see e.g. Theorem 2.1 in Newey and McFadden (1994). The second main requirement is uniform convergence of a feasible "sample" criterion function $\hat{Q}_n(\theta)$ to $Q_0(\theta)$ under the asymptotic sequence in question. Under regularity conditions, the limits of the population expectations of objects entering such a criterion $Q_0(\theta)$ under many-player sequences correspond to expectations under the limiting distributions $f(y_{mi}|\mathcal{F}_\infty)$.

for the limiting game. We consider inference and estimation for the finite player game based on the sample moments

$$\hat{m}_{nM}(\theta) := \frac{1}{nM}\sum_{m=1}^{M}\sum_{i=1}^{n} m(y_{mi,n}, \hat{G}_{mn}; \theta)$$

In light of de Finetti's Theorem (and Theorem 3.1 in Kallenberg (2005), respectively), in the limit there is no loss of information by restricting our attention to moments of the type-action characters alone and not exploiting cross-player variation for estimation or inference. Rather, exchangeability of players implies that for large games any information about structural parameters can only be recovered from the (random) marginal distributions $f_m(y_{mi}|\mathcal{F}_\infty)$, either from a given realization for the mth market, or from observed variation of that distribution across markets.[17] For identification results that rely entirely on the realized marginal distributions, consistency of an estimator or test typically requires that $n \to \infty$, whereas if identification is based on variation in $f_m(y_{mi}|\mathcal{F}_\infty)$ across markets, we may require in addition that the number of markets grows large as well, $M \to \infty$.

We next discuss a few examples to illustrate how moment conditions of this type arise naturally in empirical applications. Since identification analysis is not the primary objective of this paper, the following examples are not meant to be fully general or exhaustive, but to illustrate the potential of our asymptotic results for applications.

Parametric Estimation. Consider the following additively separable model with constant coefficients

$$u(s^{(j)}, \sigma_{m,-i}, t_{mi}; \theta) := \mu_j(x_{mi}; \theta) + \delta_j(x_{mi}; \theta)' G_n(s; \sigma_{m,-i}) + \varepsilon_{ij} \quad j = 1, \dots, p$$

where the unobserved payoff shifters $\varepsilon_{mi} = (\varepsilon_{mi1}, \dots, \varepsilon_{mip})'$ are i.i.d. across agents, and independent of x_{mi}. We also assume that the distribution of ε_{mi} and the functions $\mu_j(\cdot)$ and $\delta_j(\cdot)$ are known up to the parameter $\theta \in \mathbb{R}^K$.

For a any value of the aggregate G , we can then compute the conditional choice probabilities

$$\begin{aligned}
\Phi(s^{(j)}, x; G, \theta) &:= P(s_{mi} = s^{(j)} | x_{mi} = x, G^*_{m0} = G; \theta) \\
&= P\left(\mu_j(x;\theta) + \delta_j(x;\theta)'G + \varepsilon_{mij} \geq \max_{1 \leq k \leq p} \{\mu_k(x;\theta) + \delta_k(x;\theta)'G + \varepsilon_{mik}\}\right)
\end{aligned}$$

[17]In some cases, we may observe interaction among the same players in multiple markets, generating a panel data set that is partially exchangeable, i.e. we may treat market or player labels as irrelevant to the econometric model, but may allow for both player and market level heterogeneity that is not independent of observable characteristics. However, this extension would be beyond the scope of this paper, and is left for future research.

for actions $j = 1, \dots, p$. The first-order conditions for the maximum likelihood estimator of θ are of the form $\frac{1}{nM}\sum_{m=1}^{M}\sum_{i=1}^{n} m(y_{mi}, G_{m0}^*; \theta) = 0$, where

$$m(y_{mi}, G; \theta) := \nabla_\theta \log \Phi(s_{mi}, x_{mi}; G, \theta)$$

Identification analysis for parametric models of this form is completely analogous to the single-agent discrete choice case. One conceptual difficulty arises from the fact that a rank condition for local identification of the interaction effect also depends on the equilibrium aggregates G_m^*, which are endogenous variables in this model. This typically requires cross-market variation in G_{m0}^* which can result from players coordinating on different equilibria across different markets. However if we do not want identification to rely on assumptions regarding equilibrium selection, we need the type distribution $H_m(t)$ to vary across markets in a way that we can find pairs of markets m, m' for which the resulting equilibrium sets $\mathcal{G}_m^* \cap \mathcal{G}_{m'}^* = \emptyset$.

The assumption of a parametric model for the systematic parts of payoffs and the distribution of unobserved payoff shifters can also be relaxed in many settings. For example Matzkin (1992) gives conditions for nonparametric identification of the binary choice model which can be applied to the "competitive limit" in a manner that is completely analogous to the parametric case. For the binary action case, it is also possible to relax the assumption of independence of x_{mi} and ε_{mi} and derive moment conditions identifying the parameter from a conditional median restriction for ε_{mi} given x_{mi} using arguments similar to Manski (1975)'s maximum score estimator for the single-agent binary choice problem.

Nonparametric Sign Tests. Alternatively, the researcher may be interested in qualitative aspects of strategic interaction without specifying a parametric model for payoffs. De Paula and Tang (2012) propose inference regarding the sign of the interaction effect for the private-types case. We now show how to nest their procedure into our asymptotic framework which can also be used for inference in the complete information version of the game or other informational environments.

Consider the binary action game with $\mathcal{S} = \{0, 1\}$ and payoffs

$$u(1, G, t_{mi}) - u(0, G, t_{mi}) = \mu(x_{mi}) + \delta(x_{mi})G - \varepsilon_{mi}$$

where $\mu(x)$ and $\delta(x)$ are unknown functions, and ε_{mi} is independent of $x_{m1}, \dots, x_{mn}$ conditional on certain market-level variables that are also observed by the econometrician. Let $\{x_{mj}\}_{j\geq 1} := \{x_{m1}, x_{m2}, \dots\}$ denote the unordered sample (empirical distribution) of observable types (x_{mi}). Since payoff functions are symmetric up to the individual-specific value of x_{mi}, Proposition 1 in De Paula and Tang (2012) implies that the sign of the interaction effect $\delta(x_{mi})$ is equal to the sign of

$$\Gamma_i(x_{mi}) := \mathbb{E}[s_{mi}G_{m0}^*|\{x_{mj}\}_{j\geq 1}, x_{mi}] - \mathbb{E}[s_{mi}|\{x_{mj}\}_{j\geq 1}, x_{mi}]\mathbb{E}[G_{m0}^*|\{x_{mj}\}_{j\geq 1}, x_{mi}]$$

whenever $\Gamma_i(x_{mi}) \neq 0$.

Now suppose that the sign of $\delta(x)$ is constant across values $x \in \mathcal{X}_0 \subset \mathcal{X}$. Then for any function $h(x) \geq 0$ with support on a subset of $\mathcal{X}_0$ and aggregating across market with the same marginal distribution for x_{mi} we have that

$$\Gamma_h := \mathbb{E}[h(x_{mi})\Gamma_i(x_{mi})|\{x_{mj}\}_{j\geq 1}] = \mathbb{E}[h(x_{mi})s_{mi}G^*_{m0}|\{x_{mj}\}_{j\geq 1}] - \mathbb{E}[h(x_{mi})s_{mi}|\{x_{mj}\}_{j\geq 1}]\mathbb{E}[G^*_{m0}|\{x_{mj}\}_{j\geq 1}]$$

has the same sign as $\delta(\cdot)$ by the law of iterated expectations.

For any nonnegative function of observable types $h(x) \geq 0$ such that the support of $h(\cdot)$ is a subset of $\mathcal{X}_0$, we can form moments

$$m_1(y_{mi}; G) := h(x_{mi})s_{mi}G\mathbb{1}\{\{x_{mj}\}_{j\geq 1} = \{x_j\}_{j\geq 1}\}, \quad m_2(y_{mi}; G) := h(x_{mi})s_{mi}\mathbb{1}\{\{x_{mj}\}_{j\geq 1} = \{x_j\}_{j\geq 1}\}$$

Also, let $m_3 := \mathbb{1}\{\{x_{mj}\}_{j\geq 1} = \{x_j\}_{j\geq 1}\}$ and $m_4(G) := G\mathbb{1}\{\{x_{mj}\}_{j\geq 1} = \{x_j\}_{j\geq 1}\}$.

It then follows that the sign of

$$\begin{aligned}\Gamma_h(\{x_j\}_{j\geq 1}) \;:=\;& \mathbb{E}[m_1(y_{mi}; G^*_{m0})|\mathcal{F}_\infty]\mathbb{E}[m_3|\mathcal{F}_\infty] \\ & - \mathbb{E}[m_2(y_{mi}; G^*_{m0})|\mathcal{F}_\infty]\mathbb{E}[m_4(G^*_{m0})|\mathcal{F}_\infty]\end{aligned}$$

is either zero or equal to the sign of $\delta(x)$ on $\mathcal{X}_0$. Since the limiting game has the same form as the private-types game under any of the informational settings covered by our results, we can therefore combine our asymptotic results for the moment functions with the delta-rule to derive the distribution for the (finite-player) sample analog

$$\hat{\Gamma}_h := \hat{m}_1\hat{m}_3 - \hat{m}_2\hat{m}_4$$

where $\hat{m}_l := \frac{1}{M}\sum_{m=1}^M \frac{1}{n}\sum_{i=1}^n m_l(y_{mi,n}; \hat{G}_{mn})\mathbb{1}\{\{x_{mj}\}_{j\geq 1} = \{x_j\}_{j\geq 1}\}$ for $l = 1, \dots, 4$. For any information structure covered by our results, the many-player limit of the game has the same form as the game with private types. Hence the sign restriction on Γ_h holds for the limiting sequence regardless of the specification of the information structure in the finite-player game, and our asymptotic results can be used to construct asymptotically valid inference procedures.

Exclusion Restrictions. In many cases, it may be desirable to allow for unobserved market-level shocks which may have a distribution of unknown form, and may also be correlated with observable types. Following Shang and Lee (2011), consider estimation of the binary action game with payoffs

$$u(1, G, t_{mi}) - u(0, G, t_{mi}) = x'_{mi}\beta + \delta G + \xi_m - \varepsilon_{mi}$$

where the unobserved individual taste shifter ε_{mi} is an i.i.d. standard normal random variable, and ξ_m is an unobserved market level shock that may be correlated with x_{mi}. Furthermore, the econometrician observes a vector of instrumental variables z_{mi} satisfying the exclusion restriction $\mathbb{E}[\xi_m|z_{mi}] = 0$.

Shang and Lee (2011) propose a two-step procedure: if we define $\alpha_m := \delta G_m + \xi_m$, we can estimate the model $P(s_{mi} = 1|x_{mi}, G_m, \xi_m) = \Phi(x_{mi}'\beta + \alpha_m)$. The resulting score equations can be written as the sample average of the moment functions

$$m_1(y_{mi}, G_m; \theta) := \nabla_\theta \left\{ s_{mi} \log \Phi(x_{mi}'\beta + \alpha_m) + (1 - s_{mi}) \log(1 - \Phi(x_{mi}'\beta + \alpha_m)) \right\}$$

where $\theta := (\beta', \delta, \alpha_1, \dots, \alpha_M)'$, and the sample average is taken over markets m and players $i = 1, \dots, n$. For inference with regarding the interaction parameter δ, the second step uses the moment conditions

$$m_2(y_{mi}, G_m; \theta) := z_{mi}(\alpha_m - \delta G_m)$$

where the exclusion restriction for z_i implies that $\mathbb{E}\left[m_2(y_{mi}, G_{m0}^*; \theta_0) \right| \mathcal{F}_\infty] = \mathbb{E}\left[z_{mi}\xi_m | \mathcal{F}_\infty \right] = 0$.

In the presence of market-level shocks, consistent estimation typically requires that we observe a large number of markets, $M \to \infty$, however based on our results in section 5, under regularity conditions the maximum likelihood estimator for homogeneous markets is consistent for finite M provided a rank condition for identification holds. We illustrate the performance of our methods for an estimator of this type in the Monte Carlo study in section 6.

For any information structure covered by our results, the many-player limit of the game has the same form, so that we can directly apply our asymptotic results directly after stacking the two sets of moment conditions. Other modeling assumptions may result in inequality rather than equality restrictions of this form, in which case the law of large numbers and central limit theorem in the following section can be applied to the sample moments in order to derive (set) inference procedures, see e.g. Chernozhukov, Hong, and Tamer (2007).

4.2. **Qualitative Description of Main Results.** We now give a qualitative description of our main asymptotic results on estimation and inference which is made rigorous in section 5 and appendix A. We find that the sample moment satisfies a stochastic expansion of the form

$$\hat{m}_{nM}(\theta) = m_0(\theta) + (nM)^{-1/2} Z + n^{-1} B^* + R_n^*$$

where Z is a normal random variable with mean zero and random variance, that can be estimated consistently. B^* and R_n^* are approximation errors and B^* is an element of some bounded set and the remainder satisfies $nR_n^* \to 0$ almost surely. In this expansion, we do not impose any restrictions on the selection of equilibria. The limiting distribution may retain some indeterminacy regarding equilibrium selection, but we show that this does not affect strategies for identification or estimation.

One special case of interest is maximum likelihood or moment-based estimation of the structural parameter θ. Specifically suppose that given a sample of games with n players each, the parameter of interest θ is set-identified, where we denote the identified set with $\Theta_{0,n}$.[18] Our analysis in section 4.1 implies that for many cases of interest, the set $\Theta_{0,n}$ shrinks to a point θ_0 in the "competitive limit." In that case, we can apply our limiting theory directly to the estimating equations for the point estimator $\hat{\theta}_n$ and show that their (generically unique) solution approximates any point in $\Theta_{0,n}$.

We show how to account for simultaneity in players' choices by augmenting the main moments by the equilibrium conditions for the aggregate state G^*_{mn}, which may have multiple solutions. Given the convergence results from our theoretical analysis of aggregate games in section 4, the joint problem of estimation of the parameters of interest and determining the equilibria of the finite economy for any given realization of types can be approximated by a regular estimation problem. Hence we can apply standard methods to derive the joint distribution for $\hat{m}_n(\theta)$ and G, as well as bootstrap methods for bias reduction or other refinements for the augmented model.

We show how to estimate the distribution of the second term using a central limit theorem (CLT) with mixing, and consistent estimation of the (potentially random) variance matrix. The second term in the expansion is also random, but features of its distribution can in many cases be recovered by adapting re-sampling strategies like the bootstrap or jackknife to this problem. As an example we show how to correct for the second-order bias of the maximum likelihood estimator (MLE) in a discrete game of strategic complementarities using the bootstrap, which is straightforward to implement.[19] We find that with the second-order bias correction procedure described in the appendix, we can already obtain useful approximations for games as small as 10 to 15 players. These results suggest that for n large enough, using the much simpler (but misspecified) limiting model for inference leads to approximation errors that are of the same or a lower order of magnitude than the sampling error for a fixed number of markets.[20]

[18]Note that with heterogeneously distributed markets, the identification region $\Theta_{0,n}$ may also depend on the number of markets M. However, since our results are primarily about n growing large we chose not to make that dependence explicit in our main notation.

[19]Recall that for regular models, bias corrections of this type typically reduce the theoretical estimator bias to the order n^{-2}, see e.g. Rilstone, Srivastava, and Ullah (1996), and Hahn and Newey (2004) for the case of nonlinear panel models with fixed effects. For sufficiently smooth regular estimators and statistics, iteration of the bootstrap principle generally allows for bias correction to higher orders than n^{-1}, however we do not explore this further in the context of this paper.

[20]For an increasing number of markets, inference will in general only be asymptotically valid in combination with a second-order bias correction, where we also require that M grows at a rate slower than $n^{1/3}$. In this paper, we restrict our attention to the case in which M is fixed, and we leave a more systematic analysis of "large n, large M" asymptotics for future research.

The limiting results in this paper can also be used for set inference when the parameter of interest is not point-identified in the limit. For example if the identified set can be characterized by moment inequalities, bias reduction techniques may be applied directly to the moment vector of moment equalities or inequalities characterizing the identified set.

5. Asymptotic Analysis

This section develops conditional laws of large numbers and central limit theorems for the sample moments

$$\hat{m}_n(\theta) := \frac{1}{nM}\sum_{m=1}^{M}\sum_{i=1}^{n} m(y_{mi,n}; \hat{G}_{mn}, \theta)$$

Our asymptotic results in this section are conditional on the sigma algebra generated by the marginal distributions of y_{m1}, denoted by $\mathcal{F}_\infty$. In particular, the limits are also taken conditional on the distribution, but not necessarily the individual realizations of the observable characteristics x_{mi}. We can therefore allow the empirical distribution of types to be random across markets. Since y_{mi} also includes outcomes s_{mi} that are determined endogenously in the model, the distribution of y_{mi} generally depends on which equilibrium is being selected in the market. For aggregate games, Lemma 3.1 implies that under regularity conditions (i.e. Assumptions 3.1-3.3) and conditional on the distribution of t_{mi}, $\mathcal{F}_\infty$ is generated by a finite number of tail events, corresponding to the elements of the set of limiting equilibria $\mathcal{G}^*_{m0}$.

The following analysis only considers one single market, so that we can drop the m subscript without loss of generality. The results can then be applied for each $m = 1, \dots, M$ separately, so that information from different markets can be combined for estimation or inference, where we can also allow for type distributions and parameters to be market-specific. The setup considered in this section is more general than the specific case of aggregate games, but we will discuss applicability of our results mainly in terms of that specific class of games. For the asymptotic results, we let $y_{mi,n}$ denote the type action character for the ith player in the market for the game with n players, so that the type-action profiles for market m form a triangular array $(y_{m1,n}, \dots, y_{mn,n})_{n\geq 1}$.

We maintain the following regularity conditions on the moment function $m(y, G; \theta)$ in conjunction with the type distribution $H_m(x, \varepsilon)$ given in the model definition in section 2:

Assumption 5.1. ***(Uniform Integrability)*** *(i) The family $\{m(y, G; \theta) : \theta \in \Theta\}$ is a VC class of functions, and (ii) $\mathbb{E}|m((s, x_{m1}), G; \theta)|$ and $|\mathbb{E}\left[m((s, x_{m1}), G; \theta) - m((s', x_{m1}), G; \theta) | \mathcal{F}_\infty\right]|$ are continuous in G and bounded by a constant for all $s, s' \in \mathcal{S}$ and uniformly in $\theta \in \Theta$ and $G \in \mathcal{S}$.*

We can now state our first asymptotic result which is an adaptation of Birkhoff's law of large numbers.

Theorem 5.1. ***(Conditional LLN)*** *Suppose Assumptions 3.1-3.3 and 5.1 hold. Then under any assumptions on equilibrium selection, the average*

$$\frac{1}{n}\sum_{i=1}^{n} m(y_{mi,n}, \hat{G}_{mn}; \theta) \to \mathbb{E}\left[m(y_{m1}, G^*_{m0}; \theta)|\mathcal{F}_\infty\right] \text{ a.s.}$$

as $n \to \infty$ under a coupling satisfying the requirements of Definition 3.1. Moreover, convergence is uniform in θ.

Since this law of large numbers is uniform with respect to θ, it can be used to derive consistency results for θ or parameter sets based on moment restrictions as those derived in section 4.1.

For inference and confidence statements, we next develop a distribution theory that is conditional on the tail events in $\mathcal{F}_\infty$. To this end we are going to consider *mixing* convergence in distribution: We say that for a real-valued random sequence Z_n and a random variable Z convergence in distribution, $Z_n \stackrel{d}{\to} Z$, is mixing relative to the sigma-field $\mathcal{F}_\infty$ if for all events $A \in \mathcal{F}_\infty$,

$$\lim_n P(\{Z_n \leq z\}|A) = P(Z \leq z)$$

at all continuity points z of the c.d.f. of Z, see Hall and Heyde (1980). In general, conditions on the coupling to obtain a CLT with mixing will be more restrictive than what was needed for a conditional LLN. We will focus on the case of aggregate games described in section 4, and propose a fairly intuitive approach to incorporating equilibrium conditions as additional estimating equations into a derivation of the asymptotic distribution of $\hat{m}_n(\theta)$.

We derive the asymptotic distribution of $\sqrt{n}(\hat{m}_n(\theta) - m_0(\theta))$ conditional on $\mathcal{F}_\infty$ by augmenting the estimating equations by a state condition which defines equilibrium in the limiting game, but only approximates the fixed-point condition for the finite player game. Recall first that by Proposition 3.1, any Nash equilibrium G^*_{mn} in the aggregate game satisfies the fixed-point condition

$$0 \in \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\psi_n(t_{mi}; G^*_{mn})|\, w_m\right] - G^*_{mn}$$

With some abuse of notation, in the following we write $m((x', \psi)', G; \theta)$ to denote the moment function $m((x, \cdot), G; \theta)$ evaluated at a realization of s under some distribution in ψ. We can then stack the moment and fixed point conditions, and consider the joint distribution of

$$\begin{bmatrix} \hat{m}_n(G, \theta) \\ \hat{\Psi}_n(G) \end{bmatrix} := \frac{1}{n}\sum_{i=1}^{n} \begin{bmatrix} m((\psi_n(t_{mi}; G), x'_{mi})', \hat{G}_{mn}(G); \theta) \\ \mathbb{E}\left[\psi_n(t_{mi}; G)|\, w_m\right] \end{bmatrix}$$

where $\hat{G}_{mn}(G) := \frac{1}{n}\sum_{i=1}^{n} s_n(t_i; G)$, and $(s_n(t_i; G))_{i \leq n}$ are independent realizations of random variables in $\mathcal{S}$ with respective distributions $\psi^*_n(t_i; G) \in \psi_n(t_i; G)$.

Given θ, we let the random variable $G^*_{m0} := \mathbb{E}[s_1|\mathcal{F}_\infty]$ be the limiting equilibrium value of the aggregate conditional on $\mathcal{F}_\infty$. Also, let $m^*_{m0} := \mathbb{E}[m((\psi_0(t_{mi}; G^*_{m0}), x_i')', G^*_{m0}; \theta)|\mathcal{F}_\infty]$, and $\mu_{m0} := (m^*_{m0}, G^*_{m0})'$. Similarly, define $\hat{m}_{mn} := \frac{1}{n}\sum_{i=1}^n m_n(y_{mi}, \hat{G}_{mn}; \theta)$ and $\hat{\mu}_{mn} := (\hat{m}_{mn}, G^*_{mn})$.

We can then express the stacked moment conditions defining $\hat{\mu}_n$ in a more compact notation by defining the multi-valued function

$$r_n(t_{mi}; \theta, \mu) := \left[\begin{array}{c} m((\psi_n(t'_{mi}, G), x_i')', \hat{G}_{mn}(G); \theta) - m \\ \mathbb{E}\left[\psi_n(t_{mi}; G)\right| w_m] - G \end{array}\right]$$

By inspection, holding $\theta \in \Theta$ constant, $\hat{\mu}_n$ is a solution of the inclusion

$$0 \in \hat{r}_n(\theta, \hat{\mu}_n) := \frac{1}{n}\sum_{i=1}^n r_n(t_{mi}; \hat{\mu}_n)$$

Now let $\psi_0^*(t_{mi}; G) \in \psi_0(t_{mi}; G)$ be an arbitrary selection of the limiting best response correspondence, and let

$$r(t_{mi}; \theta, \mu) := \left[\begin{array}{c} m((\psi_0^*(t_{mi}; G), x_i')', G; \theta) - m \\ \mathbb{E}\left[\psi_0^*(t_{mi}; G)\right| w_m] - G \end{array}\right]$$

It follows that for a given value of θ and under the conditions of the theorem, μ_0 is a solution of

$$0 = r_0(\theta, \mu) := \mathbb{E}\left[r(t_{mi}; \theta, \mu)|\mathcal{F}_\infty\right]$$

Note that the general structure may be similar in other settings, e.g. dynamic games or matching markets, although in these classes of games the equilibrium condition will typically be infinite-dimensional.

For the derivation of the asymptotic distribution of $\hat{m}_n(\theta)$ we denote the Jacobians

$$\begin{aligned} \dot{M}_G(\theta) &:= \nabla_G \mathbb{E}\left[m(y_{mi}, G; \theta)|\mathcal{F}_\infty\right]\big|_{G=G^*_{m0}} \\ \dot{B}_G(\theta) &:= \nabla_G \mathbb{E}\left[m((\psi_0(t_{mi}; G)', x'_{mi})', G; \theta)|\mathcal{F}_\infty\right]\big|_{G=G^*_{m0}} \end{aligned}$$

and

$$\dot{\Psi}_G := \left[\begin{array}{c} \nabla_G\, P\left(\delta_{s^{(1)}} \in \psi_0(t_{mi}; G)|\mathcal{F}_\infty\right)\big|_{G=G^*_{m0}} \\ \vdots \\ \nabla_G P\left(\delta_{s^{(p)}} \in \psi_0(t_{mi}; G^*_{m0})|\mathcal{F}_\infty\right)\big|_{G=G^*_{m0}} \end{array}\right]$$

Note that $\dot{B}_G(\theta)$ and $\dot{M}_G(\theta)$ also generally depend on θ, although for the remainder we are going to consider asymptotic distributions for a fixed value of θ and will therefore suppress the argument.

Also define

$$b_{mi}(\theta) := m(y_{mi}; \theta) - m_0^* + \dot{M}_G\left(\psi_0(t_{mi}; G^*_{m0}) - \mathbb{E}\left[\psi_0(t_{mi}; G^*_{m0})\right| w_m]\right)$$

and the conditional covariance matrices

$$\begin{aligned}
\Omega_{bb'}(\theta) &:= \mathbb{E}[b_{mi}(\theta) - m_0^*)b_{mi}(\theta)'|\mathcal{F}_\infty] \\
\Omega_{\psi\psi'}(\theta) &:= \mathbb{E}\left[\left(\mathbb{E}\left[\psi_0^*(t_{mi};G_{m0}^*)\right|w_m\right] - G_{m0}^*\right)\left(\mathbb{E}\left[\psi_0^*(t_{mi};G_{m0}^*)\right|w_m\right] - G_{m0}^*\right)'\Big|\mathcal{F}_\infty\Big] \\
\Omega_{b\psi'}(\theta) &:= \mathbb{E}\left[\left(b_{mi}(\theta)\left(\mathbb{E}\left[\psi_0^*(t_{mi};G_{m0}^*)\right|w_m\right] - G_{m0}^*\right)'\Big|\mathcal{F}_\infty\right]
\end{aligned}$$

In general this variance matrix depends on the information structure of the game. In the pure private types model, $\mathbb{E}\left[\psi_0^*(t_{mi};G_{m0}^*)\right|w_m]$ is $\mathcal{F}_\infty$-measurable, so that $\Omega_{m\psi'}$ and $\Omega_{\psi\psi'}$ are both equal to zero. In the complete information model, $\psi_0^*(t_{mi};G_{m0}^*)$ is a.s. w_m-measurable, so that the conditional expectation is equal to the random variable itself. We also let

$$\Omega(\theta) := \left[\begin{array}{cc} \Omega_{bb'}(\theta) & \Omega_{b\psi'}(\theta)' \\ \Omega_{b\psi'}(\theta) & \Omega_{\psi\psi'}(\theta) \end{array}\right]$$

where in the following we are going to suppress the argument θ with the understanding that results are for any fixed value of $\theta \in \Theta$.[21]

We now impose standard regularity conditions on the problem that ensure a joint normal asymptotic distribution for the aggregate state and the moment conditions $\hat{m}_n(\theta)$:

Assumption 5.2. *(i) The equilibrium points $G_{m0}^* \in \mathcal{G}_{m0}^*$ are interior points of $\Delta\mathcal{S}$, (ii) the class $\mathcal{M}_s := \{m((x_{mi}', s)', G; \theta) : G \in \Delta\mathcal{S}, \theta \in \Theta\}$ is Donsker with respect to the distribution of x_{mi} for each $s \in \mathcal{S}$ with a square-integrable envelope function, (iii) the eigenvalues of Ω are bounded away from zero and infinity almost surely, and (iv) for all values of $y \in \mathcal{Y}$, $m(y, G; \theta)$ is differentiable in G and θ, and the derivative $\nabla_G m(y, G; \theta)$ is uniformly bounded and continuous in G.*

The first part of this assumption ensures that the aggregate state $\hat{G}_{mn}$ is a regular parameter in the sense of Bickel, Klaassen, Ritov, and Wellner (1993) conditional on any tail event in $\mathcal{F}_\infty$. It is possible to show that this condition is satisfied e.g. if Assumption 3.3 (ii) is strengthened to hold with all eigenvalues of $\nabla_{t_{\cdot 1}}\mathbf{u}(t, G)$ bounded away from zero, and if the conditional support of $t_{\cdot 1}$ given $t_{\cdot 2}$ is equal to $\mathbb{R}^{p-1}$. Under those additional conditions, for every action $s \in \mathcal{S}$ and $G \in \Delta\mathcal{S}$ there is a positive mass of types such that s is a best response to G, so that $\Psi_{m0}(G) \in \text{int } \Delta\mathcal{S}$, and therefore every fixed point has to be in the interior of $\Delta\mathcal{S}$. Parts (ii) and (iii) of this assumption are fairly standard. Under these additional conditions, we can obtain the following CLT:

[21]Note that since marginal distributions are $\mathcal{F}_\infty$-measurable, conditioning on $\mathcal{F}_\infty$ already implies conditioning on the type distribution $H_m(x, \varepsilon)$. For a "fixed design" approach to inference that is conditional on the realized unordered sample $\{X_1, \dots, X_n\}$, the matrix $\Omega(\theta)$ can be replaced by a conditional variance matrix where in the definition of the individual components of $\Omega(\theta)$, the unconditional means m_0^* and G_{m0}^* are replaced by their respective conditional means given x_{mi}, $\mathbb{E}[m(t_{mi}, G_{m0}^*; \theta)|x_{mi}]$ and $\mathbb{E}[\psi_0^*(t_{mi}; G_{m0}^*)|x_{mi}]$. For variance estimation, these conditional means can be replaced by consistent estimators.

Theorem 5.2. *Suppose that Assumptions 3.1-3.3 and 5.2 hold. Then* $\dot{B}_G$ *and* $(I_p - \dot{\Psi}_G)^{-1}$ *are well defined, and*

$$\sqrt{n}\left(\hat{m}_n(\theta) - \mathbb{E}[m(y_{m1}, G^*_{m0}; \theta)|\mathcal{F}_\infty]\right) \stackrel{d}{\to} Z_1 + \dot{B}_G(I_p - \dot{\Psi}_G)^{-1} Z_2 \tag{5.1}$$

mixing and uniformly in θ *under a coupling satisfying the requirements of Definition 3.1, where*

$$\left[\begin{array}{c} Z_1 \\ Z_2 \end{array}\right] \sim N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} \Omega_{bb'} & \Omega_{b\psi'} \\ \Omega_{\psi b'} & \Omega_{\psi\psi'} \end{array}\right]\right)$$

This theorem is proven in the appendix. Note that despite the asymptotic conditional independence of players' actions by de Finetti's theorem, we need to correct the asymptotic variance for dependence after centering the statistic $\hat{m}_n(\theta)$ around its conditional limit. The crucial difference to the case in which the unobserved characteristics ε_{mi} are private information lies in the adjustment for the asymptotic variance of $\hat{m}_n$ for endogeneity of $\hat{G}_{mn}$ with respect to the actual realizations of the unobserved types. The correction term is different from the variance adjustment in Shang and Lee (2011)'s analysis of the private information case, where the realized value of $\hat{G}_{mn} = \frac{1}{n}\sum_{i=1}^n s_{mi}$ takes the role of a noisy measurement for its expected value in equilibrium: we already saw that in the pure private types model, $\Omega_{\psi\psi'} = 0$, so that no adjustment is needed for the variance of the moment $\hat{m}_n(\theta)$. More generally, the definition of $\Omega_{\psi\psi'}$ together with the conditional variance identity implies that the matrix $\Omega_{\psi\psi'}$ and therefore the size of the variance adjustment increase (in the positive definite matrix sense) in the amount of information that is shared among the players.

There are several important cases in which the conditional expectation of the moment function does not vary in the value of the aggregate G^*_{m0}, in which case no variance adjustment is necessary. The following result can be verified immediately from the definition of $\dot{B}_G(\theta)$:

Lemma 5.1. *Suppose the moment functions satisfy*

$$\mathbb{E}\left[m\left((\psi_0(t_{mi}; G), x'_{mi})', G; \theta\right)\middle|\, \mathcal{F}_\infty\right] = 0$$

almost surely and for all values of G *and some* θ*. Then* $\dot{B}_G(\theta) = 0$.

Examples for this property include any moment equalities characterizing the parameter θ or the score functions for the maximum likelihood estimator. It should be noted that in these cases, the premise of Lemma 5.1 typically holds only at the population parameter θ, but not alternative values, furthermore, this property is generally not robust to misspecification of the moment functions.

The result can also be easily adapted to settings in which the econometrician only observes a random subsample of $N < n$ players in the game. As before, we let q_{mi} be an indicator variable that equals one if agent i in market m is included in the sample, and zero otherwise, and $(q_{mi})_{i\leq n}$ are independent of $y_{m1},\dots,y_{mn}$. We can then define

$$\tilde{m}_N(\theta) := \frac{1}{N}\sum_{i=1}^{n} q_{mi} m(y_{mi,n}, \hat{G}_{mn};\theta)$$

and apply Theorem 5.2 to $\tilde{m}_N(\theta)$ to obtain its asymptotic distribution.

Corollary 5.1. *Suppose the assumptions of Theorem 5.2 hold, and we observe a random sample of $N \leq n$ players' type-action profiles, where $\frac{n}{N}$ is bounded and $\lim_n \frac{N}{n} \to \alpha \in [0,1]$. Then*

$$\sqrt{N}\left(\tilde{m}_N(\theta) - \mathbb{E}[m(y_{m1}, G^*_{m0};\theta)|\mathcal{F}_\infty]\right) \stackrel{d}{\to} Z_1 + \sqrt{\alpha}\dot{M}_G(I_p - \dot{\Psi}_G)^{-1} Z_2$$

mixing and uniformly in θ under a coupling satisfying the requirements of Definition 3.1, where Z_1 and Z_2 have the same properties as before.

This result also helps highlight the different roles played by the number of players in the game and sample size in the asymptotic experiment. Specifically, the variance adjustment accounting for the equilibrium condition is important only if the size of the observed sample is of the same order of magnitude as the number of agents in the market.

5.1. **Variance Estimation.** We now turn to estimation of the conditional asymptotic variance for $\hat{m}_n(\theta)$: Given a sample $y_{m1},\dots,y_{mn}$ we can estimate the Jacobians $\dot{B}_G$, $\dot{M}_G$, and $\dot{\Psi}_G$ either parametrically given the distribution of t_{mi}, or nonparametrically to obtain $\widehat{\dot{B}}_G$, $\widehat{\dot{M}}_G$, and $\widehat{\dot{\Psi}}_G$. For semi-parametric index models - including our second application of semiparametric estimation of the sign of the interaction effect - this derivative can be estimated at a root-n rate in the presence of a continuous observed covariate with a nonzero coefficient, e.g. using Powell, Stock, and Stoker (1989)'s weighted average derivative estimator with constant weights (see their Corollary 4.1), or the estimator proposed by Horowitz and Härdle (1996) if only a finite number of markets are observed, so that variation in G^*_{mn} is only discrete. We therefore state the following high-level assumption for consistent variance estimation:

Assumption 5.3. *(i) For every value of $s \in \mathcal{S}$, the first two moments of $|m((s, x'_{mi})', G;\theta)|$ are bounded by a constant for all $G \in \Delta\mathcal{S}$ $\theta \in \Theta$, (ii) there exist consistent estimators $\widehat{\dot{B}}_G$, $\widehat{\dot{M}}_G$, and $\widehat{\dot{\Psi}}_G$ for the Jacobians $\dot{B}_G$, $\dot{M}_G$, and $\dot{\Psi}_G$, respectively, and (iii) we have a consistent estimator $\hat{\psi}_{mi,n}$ for the conditional expectation $\mathbb{E}[\psi^*_0(t_{mi}; G^*_{mn})|w_m]$.*

For primitive conditions for part (ii) of this assumption, see e.g. Powell, Stock, and Stoker (1989) and Horowitz and Härdle (1996), where we can combine their consistency arguments with the conditional law of large numbers in Theorem 5.1. Note in particular that identification of the Jacobians using the arguments in Horowitz and Härdle (1996) requires that we observe at least two markets with different limiting values of the aggregate state, either due to differences in type distributions or equilibrium selection.

For part (iii), a suitable estimator $\hat{\psi}_{mi,n}$ has to account for the assumed (or estimated) information structure of the game: in the complete information case, $\psi_0^*(t_{mi}; G_{m0}^*)$ is under Assumption 3.3 a.s. unique. Hence we have that w.p.a.1, $\psi_0^*(t_{mi}; G_{mn}^*) = \delta_{s_{mi}}$, the unit vector in the direction corresponding to the action s_{mi}. If the public signal w_m only contains information about the observable types $x_1, x_2, \dots$, then a consistent parametric or nonparametric estimator for the conditional expectation of $\delta_{s_{mi}}$ given w_m can be used for $\hat{\psi}_{mi,n}$. If the public signal is partially informative about the unobserved types $\varepsilon_1, \varepsilon_2, \dots$, and we have a parametric model for the joint distribution of w_m and the unobserved types, a parametric estimator may be constructed from that model. This includes the setup in Grieco (2012) in which unobserved shocks are the sum of two independent normal random variables, $\varepsilon_{mi}+\xi_{mi}$, where ξ_{mi} is publicly observable and ε_{mi} is a private signal.

We can now define

$$\hat{v}_{mi,n}(\theta) := [m(y_{mi,n}, \hat{G}_{mn}; \theta) - \hat{m}_n(\theta)] + \widehat{\dot{M}}_G[\delta_{s_{mi}} - \hat{\psi}_{mi,n}] + \widehat{\dot{B}}_G(I_p - \widehat{\dot{\Psi}}_G)^{-1}[\hat{\psi}_{mi,n} - \hat{G}_{mn}] \quad (5.2)$$

Then a consistent estimator for the asymptotic variance is given by

$$\hat{V}_n = \hat{V}_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} \hat{v}_{mi,n}(\theta)\hat{v}_{mi,n}(\theta)'$$

We also let $A := [I_q, (I - \dot{\Psi}_G')^{-1}\dot{B}_G']'$. Our findings are summarized in the following corollary.

Corollary 5.2. *Suppose the conditions of Theorem 5.2 and furthermore Assumption 5.3 hold. Then*

$$\hat{V}_n \to A'\Omega A \text{ a.s.}$$

under a coupling satisfying the requirements of Definition 3.1. Furthermore,

$$\sqrt{n}\hat{V}_n^{-1/2}(\hat{m}_n(\theta) - \mathbb{E}[m(y_{m1}, G_{m0}^*; \theta)|\mathcal{F}_\infty]) \xrightarrow{d} N(0, I_q) \quad \text{(mixing)}$$

See the appendix for a proof. Variance estimation for statistics based on a random subsample of agents as considered in Corollary 5.1 is completely analogous where $v_{mi,n}(\theta)$ is replaced with

$$\hat{v}_{mi,N}(\theta) := [m(y_{mi,n}, \hat{G}_{mN}; \theta) - \hat{m}_N(\theta)] + \widehat{\dot{M}}_G[\delta_{s_{mi}} - \hat{\psi}_{mi,N}] + \sqrt{\frac{N}{n}}\widehat{\dot{B}}_G(I_p - \widehat{\dot{\Psi}}_G)^{-1}[\hat{\psi}_{mi,N} - \hat{G}_{mN}]$$

where the estimators for $\dot{M}_G$, $\dot{\Psi}_G$, and $\hat{\psi}_{mi,N}$ are replace with their subsample analogs, and $\hat{G}_{mN} := \frac{1}{N}\sum_{i=1}^{n} q_{mi}\hat{\psi}_{mi,N}$ is the empirical distribution of actions in the observed subsample.

6. Simulation Study

For ease of exposition, we restrict attention to the case of binary choice with social interactions, so that in the baseline case of anonymous interactions, the aggregate state G is a scalar. The unobserved characteristic $\varepsilon_{mi} \sim N(0,1)$ is i.i.d. across agents and independent of x_{mi}. The component ε_{mi} is not observed by the econometrician, but the full types $t_{mi} = (x_{mi}', \varepsilon_{mi})'$ are common knowledge among players.

Our simulated data is generated by a complete-information model in which interactions are anonymous,

$$u_{mi}(\mathbf{s}, \mathbf{t}) := s_{mi}\left(x_{mi}'\beta + \Delta_0 \frac{1}{n}\sum_{j=1}^{n} s_{mj} + \varepsilon_{mi}\right) \tag{6.1}$$

with an interaction effect $\Delta_0 \geq 0$. Note that since Δ_0 is nonnegative, the game is one with strategic complementarities. We can therefore use a tâtonnement algorithm to find the smallest and the largest equilibria (which are both in pure strategies) for any size of the market.[22]. Since actions are binary and the adaptive best-response dynamics starting at the infimum (supremum) of the strategy space are non-decreasing (nonincreasing), we can implement an algorithm that finds either extreme equilibrium in at most n steps.

We first demonstrate convergence of the best response correspondence underlying the basic argument for conditional convergence for aggregate games in section 4. In order to obtain a parsimonious simulation design that generates multiple equilibria, we model types as including both discrete and continuous components. To be specific, x_i is a discrete variable which takes values -5 or $+5$ with probability 0.2 each, and zero with probability 0.6. It is easy to verify that due to the positive sign of the interaction effect Δ_0, $\psi_n(t; G)$ is single-valued for all finite n and values of G and t. Figure 6 shows the average response correspondence $\hat{\Psi}_{mn}(G)$ for a single realization of a market with $n = 5, 20$ and 100 players, respectively together with the limiting function $\Psi_{m0}(G)$. The limiting best response mapping $\Phi_0(G)$ has exactly three fixed points, and for the realizations of the simulation draw shown in the figure, the finite-player versions of the game have three or five different equilibrium values for G_{mn}^*. In general there may be multiple equilibria supporting the same value of G_{mn}^*, and the probability that the number of distinct equilibrium distributions in the finite player game coincides with the number of fixed points of $\Psi_0(G)$ is strictly less than one.

We now turn to simulations of statistics of the type analyzed in section 5 in order to assess the quality of the asymptotic approximations. The Jacobians in the expression for the variance in equation (5.1) can be obtained if we have an estimator for the conditional marginal effects given x_{mi}, $q(x, G) := \nabla_G \mathbb{E}\left[\psi_0(t_{mi}; G)|x_{mi} = x\right]$. If the conditional distribution of ε_{mi} given x_{mi} is continuous and fixed with respect to x_{mi} with p.d.f. $h_\varepsilon(z)$, the conditional model for $\psi_0(t_{mi}; G)$ given x_{mi} has the linear index structure considered by Powell, Stock,

[22]see Theorem 8 and corollaries in Milgrom and Roberts (1990)

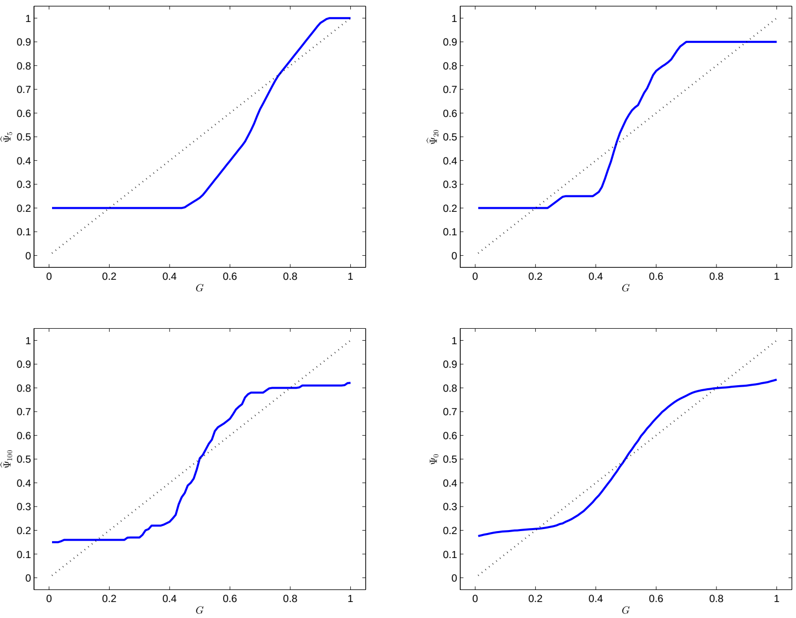

FIGURE 2. $\hat{\Psi}_{mn}(G)$ for $n = 5, 20, 100$, and $\Psi_{m0}(G)$ (bottom right)

and Stoker (1989) and Horowitz and Härdle (1996), and we can obtain consistent estimators for the index coefficients $\hat{\Delta}_n$ and $\hat{\beta}_n$, and for the conditional p.d.f. $\hat{h}_n(\cdot)$. We can then form $\hat{q}_n(x, G) := \hat{\Delta}\hat{h}_\varepsilon(x'\hat{\beta}_n + \hat{\Delta}_n G)$. Since in our setup observable types are discrete and the number of markets is finite, the conditional marginal effects are in general not point-identified without parametric assumptions on the conditional distribution of ε_{mi} given x_{mi}. We therefore use the usual parametric estimator for $q(x, G)$ based on the Probit specification, $\hat{q}(x, G) := \frac{\hat{\Delta}_n}{\sigma}\varphi\left(\frac{x'\hat{\beta}_n + \hat{\Delta}_n G}{\sigma}\right)$, where $\varphi(\cdot)$ is the standard normal p.d.f..

For the binary action case, let $d(x; \theta) := m((1, x')'; \theta) - m((0, x')'; \theta)$ so that we can express

$$\dot{M}_G = \mathbb{E}\left[d(x_{mi}; \theta)q(x_{mi}, G^*_{m0})|G^*_{m0}\right] \qquad \text{and } \dot{\Psi}_G = \mathbb{E}[q(x_{mi}, G^*_{m0})|G^*_{m0}]$$

We can then obtain estimates of the Jacobians by replacing $q(x, G)$ with its estimator $\hat{q}(x, G)$, G^*_{m0} with $\hat{G}_{mn}$, and expectations with sample averages over the observed values of x_{mi},

$$\widehat{\dot{M}}_G = \frac{1}{n}\sum_{i=1}^{n} d(x_{mi}; \theta)\hat{q}(x_{mi}, \hat{G}_{mn}) \qquad \text{and } \hat{\dot{\Psi}}_G = \frac{1}{n}\sum_{i=1}^{n} \hat{q}(x_{mi}, \hat{G}_{mn})$$

Given a consistent estimator $\hat{q}(x, G)$ for the conditional marginal effect, we can plug these estimators into the expression for $\hat{v}_{in}$ in equation (5.2) and obtain a consistent estimator for the variance matrix of $\hat{m}_n(\theta)$. Figure 6 shows kernel density approximations to the simulated p.d.f.s for G^*_{mn} and the corresponding t-ratios, where we standardize the deviations of G^*_{mn} from the respective limiting values with the estimated standard deviation based on (5.2).

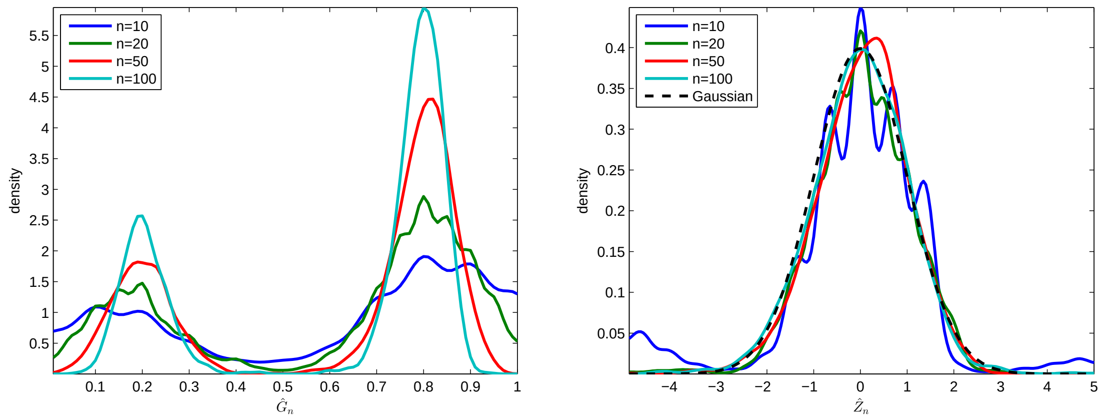

FIGURE 3. Simulated p.d.f. of $\hat{G}_{mn}$ (left panel), and its studentization $\hat{Z}_n = \sqrt{n}\frac{\hat{G}_{mn}-G^*_{m0}}{\sqrt{\hat{V}_{G,n}}}$ (right panel) for $n = 10, 20, 50, 100$.

Next we report results on the performance of estimators for the parameters β and Δ for the binary game, where we vary the number of players in each game, n, and hold the number of markets constant at $M = 20$. The specification of this model includes a constant $\beta_0 = 0.5$ and a coefficient on the scalar player characteristic x, $\beta_1 = 1$, where x was generated from a normal distribution with marginal variance equal to one and a within-market correlation coefficient equal to 0.5. In this setup, it is in general necessary to have different distributions of observable characteristics in different markets to generate sufficient variation in $\hat{G}_{mn}$ to identify the interaction parameter Δ_0, which is set equal to 1 in our simulation experiments. In the presence of multiple equilibria, we select the highest equilibrium in half the markets, and the lowest equilibrium in the remaining markets.

Table **??** reports measures of location and scale for the distribution of the maximum likelihood estimator $\hat{\theta}^{ML}$ based on the limiting model, as well as the bias corrected estimator $\hat{\theta}^{BC}$. The estimator $\hat{\theta}^{BC}$ results from the bootstrap bias reduction procedure described in the appendix. We discard bootstrap draws which result in perfect classification, or for which the maximization routine for computing the MLE does not converge after 100 iterations. Simulation results are based on 1,000 Monte Carlo replications, and the bias corrected estimator uses $S = 100$ bootstrap samples. Computation of the estimators is straightforward and was done using Matlab's standard routine for estimating generalized linear models, and obtaining the full set of simulation results reported in the tables took less than 4 hours on a standard laptop computer.

The simulation results show that the MLE based on a first-order approximation to the limiting game exhibits severe biases for small values of n, as expected, especially for the interaction parameter Δ_0. Notice also that since the number of markets is fixed at $M = 20$, the standard error for the estimated interaction parameter decreases only relatively slowly

n	$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\Delta}_0^{ML}$	$\hat{\beta}_0^{BC}$	$\hat{\beta}_1^{BC}$	$\hat{\Delta}_0^{BC}$
5	0.609	0.996	1.140	0.494	0.978	1.034
	(0.168)	(0.304)	(0.646)	(0.159)	(0.272)	(0.381)
8	0.555	1.001	1.095	0.502	1.005	0.968
	(0.150)	(0.249)	(0.587)	(0.133)	(0.223)	(0.354)
10	0.529	0.989	1.125	0.497	1.001	0.978
	(0.136)	(0.223)	(0.597)	(0.124)	(0.200)	(0.377)
15	0.513	0.989	1.116	0.502	1.009	0.975
	(0.126)	(0.180)	(0.498)	(0.111)	(0.162)	(0.337)
20	0.504	0.989	1.120	0.502	1.011	0.987
	(0.117)	(0.163)	(0.476)	(0.104)	(0.146)	(0.341)
50	0.492	0.983	1.084	0.498	0.998	1.000
	(0.079)	(0.099)	(0.303)	(0.072)	(0.093)	(0.254)
100	0.491	0.992	1.051	0.497	1.000	1.003
	(0.060)	(0.074)	(0.231)	(0.057)	(0.071)	(0.209)
200	0.498	0.998	1.020	0.501	1.002	0.995
	(0.043)	(0.053)	(0.168)	(0.042)	(0.052)	(0.160)
DGP	0.500	1.000	1.000	0.500	1.000	1.000

Table 1. Mean and standard deviation (in parentheses) of MLE and bootstrap bias corrected MLE

as n grows. We can also see that the bootstrap bias correction removes a substantial part of that bias with only minor or no increases in estimator dispersion.[23] While even after bias reduction, the estimator for the interaction parameter at $n \leq 15$ still exhibits a bias that is substantial in terms of absolute size (around 0.03), it is substantially smaller than the standard error of the estimator (around one tenth of the magnitude) even for a game with as few as $n = 5$ players. Hence for estimation, the approximation bias is not particularly large even for games with a relatively small number of players in terms of its relative contribution to the asymptotic mean-square error or other measures of estimator precision. However, for inference, biases of this magnitude will lead to severe size distortions, so that we may choose to rely on approximations of this type only for games that have a moderate to large number of players.

We also simulate the analytical standard errors implied by Theorem 5.2 and Corollary 5.2. Note that by Lemma 5.1 it is not necessary to adjust the asymptotic variance of the MLE for dependence of players' actions despite having generated the data from a complete information game. Furthermore, since the bias correction is only of the order n^{-1}, the analytical standard errors for the bias-corrected and uncorrected MLE are the same to first order. Table **??**

[23]Simulations with other parameter values and type distributions give similar results. However in cases in which we have near-violations of the regularity condition in Assumption 3.2 (e.g. if the mapping Φ_0 has two fixed points that are very close to each other), the performance of the bias reduction procedure deteriorates substantially for small values of n, as should be expected.

n	$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\Delta}_0^{ML}$	$\hat{\beta}_0^{BC}$	$\hat{\beta}_1^{BC}$	$\hat{\Delta}_0^{BC}$
5	0.172	0.290	0.811	0.172	0.290	0.811
	(0.168)	(0.304)	(0.646)	(0.159)	(0.272)	(0.381)
8	0.162	0.245	0.758	0.162	0.245	0.758
	(0.150)	(0.249)	(0.587)	(0.133)	(0.223)	(0.354)
10	0.160	0.225	0.728	0.160	0.225	0.728
	(0.136)	(0.223)	(0.597)	(0.124)	(0.200)	(0.377)
15	0.147	0.191	0.646	0.147	0.191	0.646
	(0.126)	(0.180)	(0.498)	(0.111)	(0.162)	(0.337)
20	0.135	0.169	0.575	0.135	0.169	0.575
	(0.117)	(0.163)	(0.476)	(0.104)	(0.146)	(0.341)
50	0.092	0.107	0.375	0.092	0.107	0.375
	(0.079)	(0.099)	(0.303)	(0.072)	(0.093)	(0.254)
100	0.065	0.075	0.259	0.065	0.075	0.259
	(0.060)	(0.074)	(0.231)	(0.057)	(0.071)	(0.209)
200	0.045	0.052	0.177	0.045	0.052	0.177
	(0.043)	(0.053)	(0.168)	(0.042)	(0.052)	(0.160)

Table 2. Analytical standard errors (means across simulation draws) and simulated standard deviation (in parentheses) of MLE and bootstrap bias corrected MLE

reports the mean of the analytical standard error across simulated samples together with the simulated standard deviation of the uncorrected and the bias corrected MLE, respectively, in parentheses (the latter coincide with the simulated standard deviations in Table **??**). We find that the analytical standard errors somewhat overestimate true estimator dispersion for small n, but approximate the simulated quantities as n grows large. The approximation works substantially better for the uncorrected MLE and the parameters β_0 and β_1. In practice, it may be preferable to use bootstrap or simulation methods to obtain higher-order accurate standard errors for games of small or moderate size.

In order to evaluate inference based on the asymptotic results in section 5, we also simulate null rejection frequencies for tests based on the individual coefficients. Specifically, we test the null of the "true" respective values of the parameters $\beta_0, \beta_1, \Delta_0$ used for the data generating process using t-tests based on the MLE (both corrected and uncorrected) at a nominal size of $\alpha = 0.05$. For all tests, we use the analytical standard errors and Gaussian limiting approximation suggested by Corollary 5.2. Overall, the tests based on the bias-corrected estimator rejects at a rate of less than 0.055 for all $n \geq 8$, however rejection rates for the interaction parameter Δ_0 are distorted downwards due to the upward bias of the analytical standard error for $n \leq 50$. Tests based on the uncorrected MLE overreject at a substantial rate for small to intermediate values of n.

n	$\hat{\beta}_0^{ML}$	$\hat{\beta}_1^{ML}$	$\hat{\Delta}_0^{ML}$	$\hat{\beta}_0^{BC}$	$\hat{\beta}_1^{BC}$	$\hat{\Delta}_0^{BC}$
5	0.089	0.079	0.145	0.056	0.060	0.023
8	0.061	0.081	0.122	0.038	0.053	0.011
10	0.044	0.085	0.146	0.047	0.052	0.021
15	0.050	0.069	0.097	0.039	0.035	0.012
20	0.067	0.083	0.132	0.043	0.041	0.024
50	0.066	0.070	0.081	0.046	0.043	0.024
100	0.069	0.061	0.071	0.048	0.051	0.036
200	0.052	0.062	0.063	0.046	0.052	0.049

Table 3. False rejection rates for a t-test for the true parameter values at the nominal 95% significance level

Appendix A. Bias Reduction

In this appendix, we give a description of a parametric bootstrap procedure for bias reduction which we implemented for the simulations in section 6. It is important to note that the (singleton-valued) bias correction is in general only valid if the mapping $\psi_n(t;G)$ is single-valued at all values $t = t_{m1}, \dots, t_{mn}$ and G in some neighborhood of G^*_{m0}. For the simulation experiments in section 5, the positive sign of the interaction parameter Δ_0 guarantees that this is indeed the case in that specific example. Our discussion of this specific case is meant to illustrate the usefulness of asymptotic expansions for estimation problems in which the number of players is not large enough for the first-order approximations to be precise enough. The formulation of a more general (potentially set-valued) bias correction procedure is beyond the scope of this paper and will be left for future research.[24]

Resampling as a systematic means for (higher-order) bias reduction was first proposed by Quenouille (1956) and James Tukey. Bias correction using bootstrap was proposed in Efron (1979)'s seminal paper. For a comparison between Jackknife and analytical bias corrections in the context of nonlinear panel models, see also Hahn and Newey (2004). For expositional convenience, we focus on the case of complete information, however a similar strategy can be used for other assumptions on the information structure of the game. For this part, denote the MLE for $\beta = (\theta', G_0')'$ using the full sample with $\hat{\beta}_n = (\hat{\theta}_n', \hat{G}_{mn}')'$, assuming that the parameter θ is point identified, and other regularity conditions are met.

For the bootstrap estimator of the bias term, we generate B conditional bootstrap samples, where the bth sample is of the form $\left(\tilde{t}_{mi}\right)_{i\leq n}$, where $\tilde{t}_{mi} = (x_{mi}', \tilde{\varepsilon}_{mi}')'$ and $\tilde{\varepsilon}_{mi}$ is a random draw from the distribution $H(\varepsilon)$. We then let $\tilde{\beta}_b := (\tilde{\theta}_b', \tilde{G}_b')'$ be the solution of the equations

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^n m((x_{mi}', \psi_n(\tilde{t}_{mi}; \tilde{G}_b)); \tilde{\theta}_b) \\ \tilde{G}_b &= \frac{1}{n}\sum_{i=1}^n \psi_n(\tilde{t}_{mi}; \tilde{G}_b, \hat{\theta}_n) \end{aligned}$$

where in the case of multiple roots for the equilibrium conditions, we pick the solution that is closest to $\hat{G}_{mn}$.

[24]Note that it is generally possible to form conservative bounds for the higher-order bias based on the influence functions for θ, where we take component-wise minima and maxima over the potentially non-unique best responses $s \in \psi_n(t_{mi}; G)$.

For incomplete information games, we can simply replace the second equation in the previous display with

$$\tilde{G}_b = \frac{1}{n}\sum_{i=1}^{n} \widehat{\mathbb{E}}\left[\psi_n^*(\tilde{t}_i; \tilde{G}_b, \hat{\theta}_n)\middle| w_m\right]$$

if w_m is observed by the researcher, where $\widehat{\mathbb{E}}[\cdot]$ denotes a parametric or semi-parametric estimator of the respective conditional choice probabilities for $s^{(1)}, \dots, s^{(p)}$ given w_m. If the public signal w_m contains information that is not observed by the researcher, then $\tilde{\theta}_b$ and $\tilde{G}_b$ could be obtained by solving the above equations where the unobserved components of w_m are replaced with simulation draws that are independent for $b = 1, \dots, B$, resulting in simulated public signals $\tilde{w}_{m,1}, \dots, \tilde{w}_{m,B}$.

The bootstrap bias corrected estimator for θ is then given by

$$\hat{\theta}^{BC} = 2\hat{\theta}_n - \frac{1}{B}\sum_{b=1}^{B}\tilde{\theta}_b$$

It is beyond the scope of this paper to provide an analytical derivation of the higher-order bias of the regular MLE $\hat{\theta}$, but to illustrate the basic idea behind this procedure, note that the expectation of a regular estimator for $\beta = (\theta', G_0')'$ generally admits an expansion

$$\mathbb{E}_F[\hat{\beta}] = \beta + \frac{\text{Bias}(F)}{n} + O\left(n^{-2}\right)$$

where the θ subscript indicates an expectation taken with respect to the population distribution F which may be characterized by θ and other nuisance parameters. Note that from a CLT, the $O_p\left(n^{-1/2}\right)$ term in the stochastic expansion of a regular estimator $\hat{\beta}$ has expectation equal to zero. This expansion can be shown to be valid for the parameter values in section 6, where the positive sign of the interaction effect guarantees that $\psi_n(t_{mi}; G)$ is single-valued for each individual and all values of G.[25]

The bootstrap estimator of the bias uses samples from an estimate $\hat{F}_n$ of the population distribution, where $\hat{\beta}_n = \beta(\hat{F}_n)$ is a smooth functional of the estimated distribution. Therefore, the MLE based on a bootstrap sample admits an analogous expansion

$$\mathbb{E}_{\hat{F}_n}[\hat{\beta}_n] = \hat{\beta}_n + \frac{\text{Bias}(\hat{F}_n)}{n} + O\left(n^{-2}\right)$$

Since $\hat{F}_n$ is known, we can either compute or approximate the expectation on the left hand side as an average over estimates $\tilde{\beta}_s$ obtained from bootstrap samples $b = 1, \dots, B$, and obtain an estimator of the bias term,

$$\frac{1}{n}\widehat{\text{Bias}} := \frac{1}{B}\sum_{b=1}^{B}\tilde{\theta}_b - \hat{\beta}_n = \frac{Bias(\hat{F}_n)}{n} + O\left(n^{-2}\right) + O_P\left(B^{-1/2}\right)$$

In the simulation experiments in section 5, the data generating process F is assumed to be known up to the parameter θ and an equilibrium selection mechanism, so that (first-order) consistency of $\hat{\theta}_{ML}$ implies convergence of the empirical law. Hence, if the second-order bias $\text{Bias}(F)$ is a continuous functional of F with respect to some norm, and the empirical or estimated law $\hat{F}$ converges weakly to F with respect to that same norm, the bootstrap bias corrected estimator satisfies

$$\mathbb{E}[\hat{\theta}^{BC}] = \theta + \frac{\text{Bias}(F) - \mathbb{E}[\text{Bias}(\hat{F}_n)]}{n} + O\left(n^{-2}\right) = \theta + O\left(n^{-2}\right)$$

[25]However in general $B(F)$ may be set valued, so that the bias reduction procedure described in this appendix is not universally valid. We leave the construction of (potentially set-valued) bias corrections for a more general setting for future research.

Note that in general sampling error in the equilibrium value of G^*_{mn} contributes to the higher-order bias in the MLE for θ, so that in general it is necessary to re-solve for the new equilibrium values $\tilde{G}_b$.

Appendix B. Extensions of the Baseline Model

This appendix discusses extensions of the baseline version of the aggregate game introduced in section 2. For one, we show how to allow for aggregate states that are a generalized index of players' types and actions. Furthermore, we show how to extend our results for aggregate games to cases in which interaction effects may be type-specific. We also generalize the convergence results from section 3 to the case in which the population best-response mapping $\Psi_{m0}(G)$ may be set-valued.

B.1. **Aggregate States depending on Player Types.** In the baseline version of our game-theoretic model, payoffs were assumed to depend only on the empirical distribution of other players' actions. In some applications, the relevant state variable is a more general index aggregating players' types and actions within the market. Specifically, payoffs may depend on an aggregate state variable of the form

$$G_{mn}(\sigma) := \frac{1}{n}\sum_{s\in\mathcal{S}}\sum_{j=1}^{n}\mathbb{E}\left[\sigma_{mj}(s)K(s,t_{mj},\xi_m)\right| w_m]$$

where ξ_m is a vector of market-specific state variables that takes values in some set Ξ, and $K:\mathcal{S}\times\mathcal{T}\times\Xi\to\mathbb{R}^q$ is a known function.

For example, in an entry game the entry of a "larger" firm into a market may have a larger effect on other potential entrants' profits than entry of a small competitor. If firms compete in a static Cournot oligopoly upon entering the market, then the continuation values for the entry decision depend on an index of marginal costs among entrants rather than only their number. As an empirical example, Ciliberto and Tamer (2009)'s heterogeneous competitive effects specification allows for each airline to have a different effect on all other competitors, but e.g. entry of American Airlines has the same effect on each of the other carriers.

As another example, Todd and Wolpin (2012) develop a model for education outcomes where teachers' and students' effort are complementary inputs in students' mastery of the curriculum. If the teacher is rewarded based on the average level of student knowledge in the class room at the end of the year, her optimal effort choice depends on an index aggregating students' effort levels in the classroom. Hence, students' effort choices are strategic complements in incentivizing teacher effort, and are determined in a static coordination game of complete information. While student effort e_{mi} is modeled as a continuous choice, their model includes a fixed cost for exerting any effort that is strictly greater than a minimal level $\underline{e}_{mi}$. Denoting the interior solution to the student's optimization problem after incurring the fixed cost with $e^*_{mi} > \underline{e}_{mi}$, the student's problem reduces to a discrete choice between the corner solution $\underline{e}_{mi}$ or the interior solution e^*_{mi}, and we can define the binary choice indicator $s_{mi} := \mathbb{1}\{e_{mi} > \underline{e}_{mi}\}$. Given that notation, Todd and Wolpin (2012) show that in their model that student i's optimal effort level depends only on his own type t_{mi}, teacher and classroom characteristics ξ_m, and the state variable of the form[26]

$$G_{mn}(\sigma) := \sum_{j=1}^{n} s_{mj}K(t_{mj}),$$

where $K(t_{mj})$ is a nonlinear function of student j's initial level of knowledge, variable cost of exerting effort, and marginal utility of knowledge.

[26]See equation (7) in their paper.

Using the same notation as before in the paper, we can define the expected response function

$$\tilde{\Psi}_{m0}(G) := \mathbb{E}\left[\sum_{s\in\mathcal{S}} \psi_0(t_{mi}; G, s) K(s, t_{mi}, \xi_m)\right]$$

and the aggregate response mapping

$$\hat{\Psi}_{mn}(G) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\sum_{s\in\mathcal{S}} \psi_n(t_{mi}; G, s) K(s, t_{mi}, \xi_m)\right| w_m\Bigg]$$

where $\psi_0(t; G, s)$ and $\psi_n(t; G, s)$ denote the coordinate of $\psi_0(t; G)$ and $\psi_n(t; G)$, respectively, corresponding to the pure action s. It is then straightforward to verify that any value of G^*_{mn} supported by a Nash equilibrium has to satisfy the fixed point condition $G^*_{mn} \in \hat{\Psi}_{mn}(G^*_{mn})$.

Then, if the function $K(s, t, \xi)$ is bounded, the aggregate state variable G_{mn} only takes values in a compact subset of $\mathbb{R}^q$. Hence, if the modified mapping $\Psi_{m0}(G)$ satisfies Assumption 3.2, it has only a finite number of fixed points, and the conclusions of Theorem 3.1 and Proposition 3.2 generalize directly to this extended model. Similarly, the asymptotic results of section 5 go through for this alternative model without any major modifications.

B.2. **Type-Specific Interactions.** The reference model (2.2) assumed that players' payoffs depend only on the proportion of all players that chose a given action. In this appendix we give an extension in which players' decisions may also be influenced by the proportion of agents *of a given type* that choose each action. Specifically, suppose there is a known function $z : \mathcal{T} \to \mathcal{Z}$, where $\mathcal{Z}$ is some (finite or infinite) set, such that $z(t_{mi})$ captures all type-specific information that is relevant for any of the other players' payoffs. We can then define the extended aggregate state

$$G_{mn}(s, z; \sigma) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\sigma_{mi}(s)\mathbb{1}\{z(t_{mi}) = z\}\right| w_m\right], \qquad s \in \mathcal{S}$$

In the following, we denote the space of joint distributions over $\mathcal{Z}$ and $\mathcal{S}$ with $\Delta(\mathcal{S}\times\mathcal{Z})$. Given the extended aggregate state, preferences are assumed to be of the form

$$u_i(s, \sigma_{-i}, \mathbf{t}) = u\left(s_{mi}, G_{mn}(s, \cdot; \sigma_{-i}), t_i; \theta\right) \tag{B.1}$$

Since $G_{mn} \in \Delta(\mathcal{S} \times \mathcal{Z})$, this formulation allows for much richer forms of strategic interaction between individual agents. For example the sign and strength of the interaction effect between player i and j may be stronger if the distance between $z(t_{mi})$ and $z(t_j)$ is small, or agents with certain values $z(t_{mi})$ may have a stronger strategic impact on other players than the average player.

This extension is highly relevant for empirical applications. For example, the empirical model for airline entry in Ciliberto and Tamer (2009) groups companies as large versus medium airlines, and low cost carriers, allowing competition effects on profits to vary within and across categories. In this case, the relevant aggregate state would be the number of entrants of each type of airline. We can also use type-specific interactions to accommodate models spatial interactions, where we treat player i's "location" as part of her type t_{mi}, and the asymptotic experiments consists of adding new players at existing locations in $\mathcal{T}$. In the terminology of spatial dependence, a limiting sequence of this type can be understood as "infill" asymptotics rather than relying on weak dependence in "increasing domain" asymptotics. However in that case the assumption that the aggregate state G is finite-dimensional is restrictive.

We now define the correspondences $\psi_0(t;G)$ and $\Psi_{m0}(G) := \mathbb{E}[\psi_0(t;G)|w_m]$ as in the case of no type-specific interactions. Furthermore, we let

$$\tilde{G}_{-i,n}(G,t,z,w_m;\sigma) = \frac{n}{n-1}G - \frac{1}{n-1}\left\{ \mathbb{E}[\psi_0(t_{mi};G)\mathbb{1}\{t_{mi} \neq t\}|w_m] + \sigma P(t_{mi} = t|w_m)\right\}$$

if $z(t) = z$, and $\tilde{G}_{-i,n}(G,t,z,w_m;\sigma) = G$ otherwise. Then

$$\psi_n(t;G) := \left\{\sigma \in \Delta\mathcal{S} \mid \sigma \in \psi_0(\tilde{G}_{-i,n}(G,t,w_m;\sigma),t) \text{ and } \tilde{G}_{-i,n}(G,t,z,\sigma) \in \Delta\mathcal{S}\right\}$$

Then, following the same argument as in the proof of Propostion 3.1, the set of aggregate states supported by a Bayes Nash equilibrium equals the set of fixed points of the aggregate mapping $\hat{\Psi}_{mn}(G) := \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\psi_n(t_{mi};G)\right| w_m]$.

For the remainder of this discussion, we distinguish two cases depending on whether $\mathcal{Z}$ is finite or infinite. In the first case, $\Delta(\mathcal{S}\times\mathcal{Z})$ is finite-dimensional and compact, so that all our main formal results remain valid without any major changes. In order to allow for $\mathcal{Z}$ to be infinite, we have to allow for $D := \Delta(\mathcal{S}\times\mathcal{Z})$ to be a general Banach space. Now, for a Banach space D, the mapping $\Psi_{m0} : D \to D$ is Fréchet differentiable if there exists a continuous linear map $\dot{\Psi}_{m0} : D \to D$ such that for every compact set $K \subset D$,

$$\sup_{h\in K}\sup_{G+th\in D}\left\|\frac{\Psi_{m0}(G+th) - \Psi_{m0}(G)}{t} - \dot{\Psi}_{m0}(G)(h)\right\| \to 0$$

as $t \to 0$, see van der Vaart and Wellner (1996) section 3.9. Note that for D a subset of a Euclidean space, Fréchet differentiability is equivalent to continuous differentiability of Ψ_{m0}, where for any vector $h \in D$ the linear mapping $\dot{\Psi}_{m0}(G)(h)$ corresponds to the matrix product $\nabla_G\Psi_{m0}(G)h$, and $\nabla_G\Psi_{m0}$ is the Jacobian matrix of Ψ_{m0} at G.

For the remainder of this section, we take Assumptions 3.1 and 3.3 to hold with respect to the extended aggregate $G = G(s,z;\sigma)$. Note that neither assumption required that the aggregate G be finite-dimensional. The following assumption replaces Assumption 3.2 from the main text.

Assumption B.1. ***(Regular Economy):*** *For every $\theta \in \Theta$ and population distribution $H_0 \in \mathcal{P}$, $\Psi_{m0}(G)$ is single-valued for all $G \in \Delta\mathcal{S}$ and one of the following holds:*

- (i) *The number of distributions G^*_{m0} solving $G^*_{m0} \in \Psi_{m0}(G^*_{m0})$ is finite with probability 1, and the mapping $G - \Psi_{m0}(G)$ is Fréchet differentiable at each G^*_{m0} with continuously invertible derivative $(I - \dot{\Psi}_{m0}(G^*_{m0}))$.*
- (ii) *The number of cumulation points G^*_{m0} of the sequence $(G^*_{mn})_{n\geq 1}$ is finite with probability 1, and at any cumulation point $G - \Psi_{m0}(G)$ is Fréchet differentiable at each G^*_{m0} with continuously invertible derivative $(I - \dot{\Psi}_{m0}(G^*_{m0}))$.*

where I denotes the identity. (b) Furthermore, for every $\delta > 0$ we can find $\eta > 0$ such that $d(G,G^*) > \delta$ for all fixed points G^* implies $d(\Psi_{m0}(G),G) > \eta$.

The invertibility requirement on $I - \dot{\Psi}_{m0}$ generalizes the rank condition of Assumption 3.2. Since the proof of Lemma 3.1 only used local uniqueness of equilibria from the rank condition of Assumption 3.2 together with compactness of $\Delta\mathcal{S}$, the additional requirement of a finite number of fixed points for Ψ_{m0} in part (a) is only restrictive if $\Delta(\mathcal{S}\times\mathcal{Z})$ is not compact. In contrast, when $\mathcal{Z}$ is infinite we have to verify this additional requirement separately. While to the knowledge of the author, there are no general conditions guaranteeing a finite number of fixed points in that case, there are conditions on Ψ_{m0} under which the fixed point is unique - e.g. if Ψ_{m0} is a contraction mapping, then uniqueness follows from Banach's fixed point theorem. More generally, the main theorem of Kellogg (1976) gives a weaker condition for uniqueness on the fixed point mapping, only requiring that the eigenvalues of $\dot{\Psi}_{m0}(G)$ are different from 1 at all $G \in \Delta(\mathcal{S} \times \mathcal{Z})$.

Part (b) is immediately satisfied in the case of finite $\mathcal{Z}$ since $\Delta(\mathcal{S} \times \mathcal{Z})$ is compact. If Ψ_{m0} is a contraction mapping with contraction constant $\lambda < 1$, then part (b) holds with $\eta = \frac{1}{1-\lambda}\delta$.[27] However more generally, verification of this requirement for the infinite-dimensional case is not straightforward.

Together with Assumptions 3.1 and 3.3, Assumption B.1 is sufficient to establish the conclusion of Theorem 3.1. For the construction of a coupling, note that the proof of Proposition 3.2 requires the number of equilibria to be finite, but does not use compactness or finite dimension of $\Delta\mathcal{S}$ directly. Hence, the result still applies, so that the law of large numbers in Theorem 5.1 generalizes immediately to the case of type-specific interactions:

Corollary B.1. ***(Law of Large Numbers)*** *Suppose Assumptions 3.1, 3.3, B.1, and Condition 5.1 hold. Then the average*

$$\frac{1}{n}\sum_{i=1}^{n} m(y_{mi,n}, \hat{G}_{mn}; \theta) \to \mathbb{E}\left[m(y_{m1}, G^*_{m0}; \theta)|\mathcal{F}_\infty\right] \text{ a.s.}$$

If $\mathcal{Z}$ is a finite set, it is also straightforward to extend the central limit theorem in Theorem 5.2. For $\mathcal{Z} = \{z_1, \dots, z_J\}$ we can define $\bar{\psi}_0(t, z; G) := \psi_0(t; G)\mathbb{1}\{z(t) = z\}$ and the vector $\bar{\psi}_0(t; G) = \left(\bar{\psi}_0(t, z_1; G)', \dots, \bar{\psi}_0(t, z_J; G)'\right)'$. For an arbitrary selection $\bar{\psi}^*_0(t_{mi}; G) \in \bar{\psi}_0(t_{mi}; G)$, we can define $\bar{\Omega}$ as the conditional covariance matrix of $\left(m_0(t_{mi}, G; \theta)', \mathbb{E}[\bar{\psi}^*_0(t_{mi}; G)'|w_m]\right)'$ given $\mathcal{F}_\infty$. Given this notation, we can state the asymptotic distribution of the moment vector $\hat{m}_n(\theta)$ with type-specific interactions:

Corollary B.2. ***(Asymptotic Distribution)*** *Suppose Assumptions 3.1, 3.3, 5.2, and B.1 hold, and that $\mathcal{Z}$ is finite. Then the conclusion of Theorem 5.2 holds, where the matrices $\dot{M}_G$ and $\dot{\Psi}_G$ are taken to be derivatives with respect to the extended aggregate state, and the covariance matrix Ω is given by $\bar{\Omega}$.*

We do not derive the asymptotic distribution for the case of an infinite set $\mathcal{Z}$ which has the structure of a semiparametric Z-estimation problem. However, it is important to note that the proof of Theorem 5.2 relies on the existence of a coupling from Proposition 3.2, but does not require compactness or finite dimension of $\Delta\mathcal{S}$ otherwise. Given the coupling, Theorem 3.3.1 and Lemma 3.3.5 in van der Vaart and Wellner (1996) continue to apply under Assumption B.1 even if the parameter G is not finite-dimensional. A fully rigorous treatment of the case of an infinite set $\mathcal{Z}$ requires additional notation and regularity conditions and is beyond the scope of this paper.

B.3. **Convergence if $\Psi_{m0}(G)$ is Set-Valued.** This subsection discusses the case in which the limiting equilibrium mapping is set-valued, which will in general be the case if the distribution of types is not continuous. Specifically, we allow the expected best response mapping

$$\Psi_{m0}(G) := \mathbb{E}\left[\psi_0(t_{mi}; G)\right] \tag{B.2}$$

to be a correspondence $\Psi_{m0} : \Delta\mathcal{S} \rightrightarrows \Delta\mathcal{S}$. Specifically, the expectation operator $\mathbb{E}[\cdot]$ is taken to denote the (Aumann) selection expectation of the random set $\psi_0(t_{mi}; G) \subset \Delta\mathcal{S}$ and $t_{mi} \sim H_m(t)$. The Aumann selection expectation of a closed random set $X : \Omega \to 2^{\mathcal{X}}$ is defined as

$$\mathbb{E}[X] := \text{closure}\{\mathbb{E}[\xi] : \xi \in \text{Sel}(X)\}$$

[27]Suppose that G^* is the (unique) fixed point of Ψ_{m0}. Then for any $G \in \Delta(\mathcal{S} \times \mathcal{Z})$,

$$\|G-G^*\| = \|G-\Psi_{m0}(G)+\Psi_{m0}(G)-G^*\| \le \|G-\Psi_{m0}(G)\|+\|\Psi_{m0}(G)-\Psi_{m0}(G^*)\| \le \|G-\Psi_{m0}(G)\|+\lambda\|G-G^*\|$$

where we use the triangle inequality and that G^* is a fixed point of Ψ_{m0}. Hence, $\|G - \Psi_{m0}(G)\| \ge (1 - \lambda)\|G - G^*\|$ for any value of G which establishes the claim.

where $\text{Sel}(X)$ denotes the set of measurable selections $\xi(\omega) \in X(\omega)$ such that $\mathbb{E}\|\xi\|_1 < \infty$. See Molchanov (2005) for a full exposition.

In order to generalize Assumption 3.2 to set-valued mappings we first have to define a graphical derivative of a correspondence: Recall that the *graph* of a correspondence $\Phi : \mathcal{X} \rightrightarrows \mathcal{Y}$ is the set $\text{gph}\,\Phi := \{(x, y) \in \mathcal{X} \times \mathcal{Y} : y \in \Phi(x)\}$. The *contingent derivative* of the correspondence Φ at $(x_0', y_0)' \in \text{gph}\,\Phi$ is a set-valued mapping $D\Phi(x_0, y_0) : \mathcal{X} \rightrightarrows \mathcal{Y}$ such that for any $u \in \mathcal{X}$

$$v \in D\Phi(x, y)(u) \Leftrightarrow \liminf_{h\downarrow 0, u'\to u} d\left(v, \frac{\Phi(x_0 + hu') - y}{h}\right)$$

where $d(a, B)$ is taken to be the distance of a point a to a set B.[28] Note that if the correspondence Φ is single-valued and differentiable, the contingent derivative is also single-valued and equal to the derivative of the function $\Phi(x)$. The contingent derivative of Φ is *surjective* at x_0 if the range of $D\Phi(x_0, y_0)$ is equal to $\mathcal{Y}$. In the following, let the mapping $\Phi_0(G) := \Psi_{m0}(G) - G$.

Assumption B.2. ***(Regular Economy):*** *(a) For every $\theta \in \Theta$ and population distribution $H_0 \in \mathcal{P}$, one of the following holds:*

(i) *The contingent derivative of $\Phi_0(G)$ is lower semi-continuous and surjective for all G, and satisfies*

$$0 \in D\Phi_0(G^*_{m0}, 0)(u) \text{ if and only if } u = 0$$

at every distribution G^*_{m0} solving $G^*_{m0} \in \Psi_{m0}(G^*_{m0})$.

(ii) *If G^*_{m0} is a cumulation point of the sequence $(G^*_{mn})_{n\geq 1}$, then with probability 1, the contingent derivative of $\Phi_0(G^*_{m0})$ is lower semi-continuous and surjective and satisfies*

$$0 \in D\Phi_0(G^*_{m0}, 0)(u) \text{ if and only if } u = 0$$

(b) Furthermore, for every $\delta > 0$ we can find $\eta > 0$ such that for the convex hull $K(G, \eta) := \text{conv}\left(\bigcup_{d(G',G)<\eta} \Psi_{m0}(G')\right)$, we have

$$\sup_{\Psi\in K(G,\eta)} d_H\left(\Psi, \bigcup_{d(G',G)<\eta} \Psi_{m0}(G')\right) < \delta \text{ for all } G \in \Delta\mathcal{S}.$$

This is a generalization of Assumption 3.2 in the main text: it is straightforward to show that for a single-valued, differentiable function $\Psi_{m0}(G)$, the rank condition in Assumption 3.2 part (i) is sufficient for the generalized rank condition in part (i) of Assumption B.2. Furthermore, part (b) holds for any continuous function $\Psi_{m0}(G)$. Therefore, Assumption 3.2 implies Assumption B.2.

Part (b) of Assumption B.2 requires that the local range of Ψ_{m0} can be approximated by a convex set. Note that for $\eta = 0$ the Hausdorff distance between $\Psi_{m0}(G)$ and its convex hull is zero because Ψ_{m0} is convex-valued. It is also straightforward to verify that for $p = 2$, this condition is met if the correspondence $\Psi_{m0}(G)$ is continuous everywhere except at finitely many values of G.[29]

In order to understand the transversality condition for the case of a limiting correspondence Ψ_{m0}, consider the case $\mathcal{S} = \{0, 1\}$, so that $G \in [0, 1]$. If there is a fixed point G^* such that $\Psi_{m0}(G^*)$ is a convex set with nonempty interior, then the second part of Assumption 3.2 (ii) is violated if and only if there exists sequences $G_n \neq G^*$ with $\lim_n G_n = G^*$ and $\Psi_n \in \Psi_{m0}(G_n)$ such that $\lim_n \frac{\Psi_n - G^*}{|G_n - G^*|} = 0$. The regularity condition may fail e.g. if for all G in a one-sided neighborhood of G^*, $\Psi_{m0}(G)$ is set valued and contains G^*, or if in that one-sided neighborhood $\Psi_{m0}(G)$ is single valued with $\lim_{G_n \to G^*_{m0}} \Psi_{m0}(G_n) = \Psi_{m0}(G^*_{m0})$, and the one-sided

[28]See Definition 5.1.1 and Proposition 5.1.4 in Aubin and Frankowska (1990)

[29]In fact, lower semi-continuity at all except finitely many values of G is sufficient.

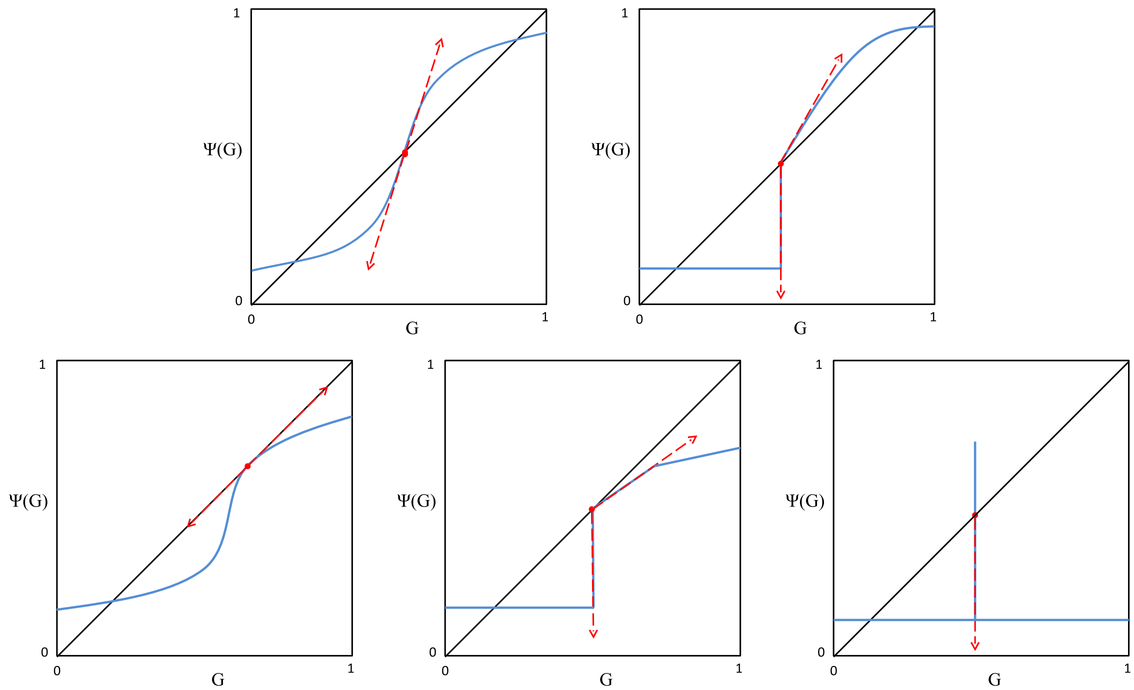

FIGURE 4. Schematic illustration of Assumption 3.2. The bottom three panels are examples of a failure of the transversality condition (left) and/or the surjectivity condition (all three panels).

derivative of $\Psi_{m0}(G^*_{m0}) = I_{p-1}$, the $(p-1)$ dimensional identity matrix. See also figure B.3 for graphical examples.

As in the main main text we say that if a fixed point $G^*_{m0} = \Psi_{m0}(G^*_{m0})$ satisfies the generalized rank condition in part (i) of Assumption B.2, it is *regular*, and we define the set of regular fixed points

$$\mathcal{G}^*_{m0} := \{G^*_{m0} : G^*_{m0} = \Psi_{m0}(G^*_{m0}) \text{ and } G^*_{m0} \text{ is a regular point of } \Psi_{m0}(G)\}$$

We can now strengthen Lemma 3.1 and Theorem 3.1 from the main text to hold under the weaker conditions in Assumption B.2.

Lemma B.1. *Under Assumption B.2 (a) (i), the number of fixed points solving $G^*_{m0} \in \Psi_{m0}(G^*_{m0})$ is finite. Furthermore, under either alternative of Assumption B.2 (a), the cardinality of $\mathcal{G}^*_{m0}$ is finite.*

Theorem B.1. *Suppose Assumptions 3.1 and B.2 hold, and that G^*_{mn} is a sequence of empirical distributions solving $G^*_{mn} \in \hat{\Psi}_{mn}(G^*_{mn})$. Then for $\mathcal{G}^*_{m0} := \{G^*_{m0} \in \Delta\mathcal{S} : G^*_{m0} \in \Psi_{m0}(G^*_{m0})\}$ and as $n \to \infty$, we have that (a) $d(G^*_{mn}, \mathcal{G}^*_{m0}) \stackrel{a.s.}{\to} 0$, and (b) with probability approaching 1, for every $G^*_{m0} \in \mathcal{G}^*_{m0}$ and every neighborhood $B(G^*_{m0})$ of G^*_{m0} we can find $\tilde{G}_n \in B(G^*_{m0})$ such that $\tilde{G}_{mn} \in \hat{\Psi}_{mn}(\tilde{G}_{mn})$.*

For the technical arguments, we refer to the proofs of Lemma 3.1 and Theorem 3.1 in the appendix, which are entirely based on Assumption B.2 rather than its stronger version in the main text.

APPENDIX C. PROOFS FOR RESULTS FROM SECTION 3

Proof of Proposition 3.1. Suppose that $G^* \notin \hat{\Psi}_{mn}(G^*)$. Then by definition of $\hat{\Psi}_{mn}$, there exists at least one player i such that $\sigma_{mi}(t_{mi}) \notin \psi_n(G^*, t_{mi})$, so that σ_{mi} is not a best response to any profile $\sigma_{m,-i}$

satisfying $\frac{1}{n}\sum_{j=1}^n \mathbb{E}[\sigma_j|w_m] = G^*$. Conversely, if $G^* \in \hat{\Psi}_{mn}(G^*)$, by the definition of Minkowski summation and selection expectations, we can find $\sigma_{m1}, \dots, \sigma_{mn}$ such that $\sigma_{mi} \in \psi_n(G^*, t_{mi})$ for all $i = 1, \dots, n$ and $\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\sigma_{mi}|w_m] = G^*$, so that G^* is indeed generated by a Nash equilibrium. Existence of an equilibrium distribution G^*_{mn} for finite n follows from an existence theorem for Bayes Nash equilibria in finite games, see e.g. Theorem 1 in Milgrom and Weber (1985), which includes Nash equilibria in mixed strategies for the complete information game as a special case $\square$

Proof of Lemma 3.1: We prove this theorem using Assumption B.2 rather than Assumption 3.2, noting that the latter is sufficient for the former as explained in Appendix B. Since the game is finite with $|\mathcal{S}| = p$, the space of distributions $\Delta\mathcal{S}$ is the $(p-1)$ dimensional probability simplex, and therefore compact. Under Assumption 3.2 (a) part (i), each equilibrium G^* is locally unique by Proposition 5.4.8 in Aubin and Frankowska (1990). We can use these open neighborhoods of the equilibrium points to cover $\Delta\mathcal{S}$, and by compactness, there exists a finite subcover $\mathcal{N}_1, \dots, \mathcal{N}_J$, say. By construction, each set $\mathcal{N}_j$, $j = 1, \dots, J$ contains at most one equilibrium point, so that the number of equilibria is at most J, and therefore finite. The second statement follows immediately, noting that by definition the set $\mathcal{G}^*_{m0}$ only includes the regular fixed points of Ψ_{m0} $\square$

Proof of Theorem 3.1. We prove this theorem using Assumption B.2 rather than Assumption 3.2, noting that the latter is sufficient for the former as explained in Appendix B.

We start by establishing claim (a), where the proof proceeds in the following steps: we first construct a convex-valued correspondence $\tilde{\Psi}_n(G)$ and show uniform convergence in probability of the convex hull of $\hat{\Psi}_n(G)$ to $\tilde{\Psi}_{mn}(G)$. We then show graphical convergence of the correspondence $\tilde{\Psi}_{mn}(G)$ to $\Psi_{m0}(G)$. Finally, we conclude that any sequence of solutions $\hat{G}^*_{mn}$ of the fixed point inclusion $\hat{G}^*_{mn} \in \hat{\Psi}_{mn}(\hat{G}^*_n)$ approaches the set $\mathcal{G}^*_{m0}$ with probability one.

Recall that the correspondence

$$\hat{\Psi}_n(G) := \frac{1}{n}\bigoplus_{i=1}^n \mathbb{E}\left[\psi_n(t_{mi}; G)\right| w_m]$$

where "$\bigoplus$" denotes the Minkowski sum over the Aumann selection expectations of the sets $\psi_n(t_{mi}; G)$ for $i = 1, \dots, n$. Also let

$$\tilde{\Psi}_n(G) := \mathbb{E}\left[\psi_n(t_{mi}; G)\right]$$

where the operator $\mathbb{E}[\cdot]$ denotes the (Aumann) selection expectation with respect to the random type $t_{mi} \sim H_0$. Note that from standard properties of Aumann expectations, $\tilde{\Psi}_{mn}(G)$ is convex at every value $G \in \Delta\mathcal{S}$.

Uniform Convergence. We will now show almost sure uniform convergence of the convex hull of $\hat{\Psi}_n(G)$ to $\tilde{\Psi}_{mn}(G)$. In order to accommodate any alternative assumptions regarding the informational content of the public signal w_m, we will first show convergence for the correspondence

$$\hat{\Psi}^*_{mn}(G) := \frac{1}{n}\bigoplus_{i=1}^n \psi_n(t_{mi}; G)$$

which is equal to $\hat{\Psi}_{mn}(G)$ in the case of complete information. We then argue that convergence for $\hat{\Psi}_{mn}(G)$ follows by the law of iterated expectations under any alternative assumptions regarding w_m.

In the following, we denote the support function of the set $\hat{\Psi}^*_{mn}(G)$ by

$$\hat{h}^*_n(v;G) := \inf_{G' \in \hat{\Psi}_{mn}(G)^*} \langle v, G' \rangle$$

for any v on the $(p-1)$ dimensional unit sphere in $\mathbb{R}^p$, and we also let

$$\tilde{h}_n(v;G) := \inf_{G' \in \tilde{\Psi}_{mn}(G)} \langle v, G' \rangle$$

Since for the Minkowski sum of two convex sets K_1 and K_2, the support function satisfies $h(a_1K_1 \oplus b_2K_2, v) = a_1 h(K_1, v) + b_2 h(K_2, v)$, we can rewrite

$$\hat{h}^*_n(v;G) = \frac{1}{n}\sum_{i=1}^{n} \varrho_n(v;G,t_{mi})$$

where at a given value of G and type t,

$$\varrho_n(v;G,t) := \inf_{G' \in \psi_n(t;G)} \langle v, G' \rangle = \inf_{G' \in \text{conv } \psi_n(t;G)} \langle v, G' \rangle$$

is the support function of the set $\psi_n(t;G)$.

In order to show that $\hat{h}_n$ converges to $\tilde{h}_n$ uniformly, i.e.

$$\sup_{v,G} \left| \hat{h}^*_n(v;G) - \tilde{h}_n(v;G) \right| \stackrel{a.s.}{\to} 0$$

note first that the sets conv $\psi_n(t_{mi};G)$ are convex polytopes whose vertices are equal to one or zero, yielding at most 2^p different sets for conv $\psi_n(t_{mi};G)$. Hence we can represent the support function $\varrho_n(v;G,t)$ as a simple function defined on sets that are p fold intersections of the sets $A(s,G)$ and their complements. Since the VC property is preserved under complements and finite intersections, Assumption 3.1 implies that $\varrho_n(v;G,t)$ is a VC class of functions indexed by (v,G) for every n with a VC index that does not depend of n.

Now by the same argument as in the proof of Theorem 3.1.21 in Molchanov (2005), almost sure convergence of the support functions implies almost sure convergence for the set conv $\left(\frac{1}{n}\bigoplus_{i=1}^n \psi_n(t_{mi};G)\right)$ to $\mathbb{E}[\psi_n(t_{mi};G)]$ with respect to the Hausdorff metric, where convergence is uniform in G by boundedness of $\varrho_n(v;G,t)$ and the VC property of the support functions for $\psi_n(t_{mi};G)$. Replacing $\psi_n(t_{mi};G)$ in the argument with its conditional expectation $\tilde{\psi}_n(w_m;G) := \mathbb{E}[\psi_n(t_{mi};G)|w_m]$, this implies that conv $\left(\frac{1}{n}\bigoplus_{i=1}^n \mathbb{E}[\psi_n(t_{mi};G)|w_m]\right)$ converges almost surely to

$$\mathbb{E}[\tilde{\psi}_n(w_m;G)] = \mathbb{E}[\mathbb{E}[\psi_n(t_{mi};G)|w_m]] = \mathbb{E}[\psi_n(t_{mi};G)]$$

where the last step follows from the law of iterated expectations. In particular, we have that

$$d_H\left(\text{conv}(\hat{\Psi}_{mn}(G)), \tilde{\Psi}_{mn}(G)\right) \stackrel{a.s.}{\to} 0$$

uniformly in G.

Graphical Convergence. Now fix $\delta > 0$, and for a set $A \subset \Delta\mathcal{S} \times \Delta\mathcal{S}$, let $A^\delta := \left\{x \in (\Delta\mathcal{S})^2 : d(x,A) \le \delta\right\}$ denote the δ-expansion of A with respect to $(\Delta\mathcal{S})^2$. In order to establish that $\tilde{\Psi}_{mn}(G) = \mathbb{E}\left[\psi_n(t;G)\right]$ converges to $\Psi_{m0}(G) = \mathbb{E}\left[\psi_0(t;G)\right]$ graphically, we show that for n large enough the two inclusions $\text{gph}\tilde{\Psi}_{mn} \subset (\text{gph}\Psi_{m0})^\delta$ and $\text{gph}\Psi_{m0} \subset \left(\text{gph}\tilde{\Psi}_{mn}\right)^\delta$ hold:

In order to verify the first inclusion, consider a point $(G_1', \Gamma')'$ such that $\Gamma \in \tilde{\Psi}_{mn}(G_1)$, where $\Gamma = (\Gamma(s^{(1)}), \dots, \Gamma(s^{(p)})) \in \Delta\mathcal{S}$. From the definition of $\tilde{\Psi}_{mn}$, it follows that there exist $G_l(s), G_u(s)$ such that $P(t \in A(s, G_u(s))) \ge \Gamma(s)$ and $P(t \notin A(s, G_l(s))) \ge 1 - \Gamma(s)$, where $d(G_1, G_l(s)) \le \frac{1}{n}$ and $d(G_1, G_u(s)) < \frac{1}{n}$

for all s. In particular, we have that $\Gamma(s) \in \mathrm{conv}\left(\bigcup_{G':d(G,G')\leq \frac{1}{n}} \Psi_{m0}(G')\right)$. Now by Assumption B.2 (b), we can find $\eta > 0$ such that $d\left(\Gamma, \mathrm{conv}\left(\bigcup_{d(G,G_1)<\eta} \Psi_{m0}(G)\right)\right) < \delta/3$, w.l.o.g. $\eta < \delta/3$. Hence, for $n \geq 1/\eta$, we also have $d\left(\Psi(s), \bigcup_{G':d(G,G')\leq\eta} \Psi_{m0}(G')\right) < \delta/3$, so that there exists G_2 such that $d(G_2, G_1) < \eta < \delta/3$ and $d(\Gamma, \Psi_{m0}(G_2)) < \delta/3$. Hence, for $n \geq 1/\eta$ we have $d((G_1', \Psi_1'), \mathrm{gph}(\Psi_{m0})) < \delta$ for all $(G_1', \Gamma')' \in \mathrm{gph}(\tilde{\Psi}_{mn})$, which establishes the first inclusion.

For the second inclusion, consider a point $(G_1', \Gamma')'$ such that $\Gamma \in \Psi_{m0}(G_1)$ and the extremal points Γ_p^* of $\Psi_{m0}(G_1)$ which are of the form $\Gamma_p^* = P(t \in A(s^{(p)}, G_1))$ for $p = 1, \dots, S-1$, and $\sum_{q\neq p} \Gamma_q^* = 1 - P(t \in A(s^{(p)}, G_1))$. It is immediate that the values $\Psi \in \Psi_{m0}(G_1)$ are convex combinations of Γ_p^*, $p = 1, \dots, S-1$, in particular $\Gamma = \sum_{q=1}^p \lambda_q \Gamma_q^*$ where $\lambda_q \geq 0$ and $\sum_{q=1}^p \lambda_q = 1$. From the construction of $\tilde{\Psi}_{mn}$, it can now be seen that for n large enough, $\Gamma \in \left(\tilde{\Psi}_{mn}(G_2)\right)^{\delta/3}$, where $G_2 := (nG_1 - \lambda)/(n-1)$. Hence, if n is in addition larger than $2/\delta$, we have $d((G_1', \Gamma'), \mathrm{gph}(\tilde{\Psi}_{mn})) < \delta$, establishing the second inclusion.

Convergence to Limiting Points. Given $\delta > 0$ we define

$$\eta := \inf \left\{ d(G, \Psi_{m0}(G)) \,|G \in \Delta\mathcal{S} : d(G, \mathcal{G}_{m0}^*) \geq \delta \right\}$$

By Lemma 3.1, the number of elements in $\mathcal{G}_{m0}^*$ is finite, so that since $\Delta\mathcal{S}$ is compact and $d_H(G, \Psi_{m0}(G))$ is lower-semi-continuous in G, we have that $\eta > 0$, where existence of the minimum follows e.g. by Theorem 1.9 in Rockafellar and Wets (1998).

Combining uniform convergence of $\hat{\Psi}_n$ to $\tilde{\Psi}_n$ in probability, and graphical convergence of $\tilde{\Psi}_n$ to Ψ_{m0} via the triangle inequality, we can choose n large enough such that the probability of $d_H(\mathrm{gph}(\mathrm{conv}(\hat{\Psi}_n)), \mathrm{gph}\Psi_{m0}) < \min\{\delta, \eta\}$ is arbitrarily close to one. Hence the probability that there exists a fixed point $\tilde{G} \in \mathrm{conv}(\hat{\Psi}_n(\tilde{G}))$ with $d(\tilde{G}, \mathcal{G}_0) > \delta$ converges to zero as n increases.

Finally, since $\hat{G}_{mn} \in \hat{\Psi}_{mn}(\hat{G}_{mn})$ implies $\hat{G}_{mn} \in \mathrm{conv}(\hat{\Psi}_n(\hat{G}_{mn}))$, we have that $d(\hat{G}_{mn}, \mathcal{G}_{m0}^*) < \delta$ with probability approaching one, which establishes assertion (a).

Achievability of Limiting Equilibria. We now turn to the proof of part (b) of Theorem 3.1. We proceed by showing that for every regular equilibrium point $G_{m0}^* \in \mathcal{G}_{m0}^*$, we can find a neighborhood of G_{m0}^* such that $\Psi_{m0}(G)$ is an inward map on that neighborhood. Then we show that by graphical convergence, $\hat{\Psi}_{mn}(G)$ is also an inward map on all such neighborhoods with probability approaching one. We then apply a fixed-point theorem for inward maps to a transformation of the problem and conclude that with probability approaching 1, local solutions exist near every $G_{m0}^* \in \mathcal{G}_{m0}^*$.

Construction of Neighborhoods. Let $\Gamma_0(G) := \Psi_{m0}(G) - G$ be defined as in the main text. In the following, we say that Ψ_{m0} is a (possibly multi-valued) *inward map* on a convex set $K \subset \Delta\mathcal{S}$, if for all $G \in K$, $\Psi_{m0}(G) \cap (G + T_K(G)) \neq \emptyset$, where $T_K(G)$ is the tangent cone of K at G. We will now show that for every $G_{0,j}^* \in \mathcal{G}_{m0}^*$, $j = 1, \dots, |\mathcal{G}_{m0}^*|$, and neighborhood $U(G_{0,j}^*)$ of $G_{0,j}^*$, we can find a compact convex set $\bar{K}_j \subset U(G_{0,j}^*)$ containing an open set around $G_{0,j}^*$, and such that Ψ_{m0} restricted to $\bar{K}_j$ is an inward map.

To this end, note first that for that every vector $u \neq 0$ and $v \in D\Gamma_0(G_{0,j}^*, G_{0,j}^*)(u)$, we have that the inner product $u'v \neq 0$, where without loss of generality we assume $u'v < 0$. Specifically, we can choose a diagonal matrix B with elements $B_{jj} \in \{-1, 1\}$, $j = 1, \dots, p-1$, and rewrite the inclusion as

$$G_{0,j}^* - B(G - G_{0,j}^*) \in \Psi_{m0}(G_{0,j}^* - B(G - G_{0,j}^*))$$

for any such B, or equivalently,

$$G = G_{0,j}^* + (G - G_{0,j}^*) \in \Psi_{m0}(G_{0,j}^* - B(G - G_{0,j}^*)) + (I - B)(G - G_{0,j}^*) =: \tilde{\Psi}_B(G)$$

Now if we let $\Gamma_B(G) := \tilde{\Psi}_B(G) - G$, by Assumption B.2 (a), the contingent derivative of Γ_B is surjective, and we can choose B such that there exists a vector $v_B \in \text{gph } D\Gamma_B(G^*_{0,j}, G^*_{0,j})(u)$, the inner product $u' \cdot v_B \leq 0$. The second part of Assumption B.2 (i) implies that this inequality is strict, and has the same sign for any other $v'_B \in \text{gph } D\Gamma_B(G^*_{0,j}, G^*_{0,j})(u)$ by convexity of the contingent derivative.

We will prove this claim by contradiction: Suppose that for a sequence $h_n \to 0$ we can construct a sequence of closed balls K_n of radius h_n around the fixed point $G^*_{0,j}$ such that for every n there exists G_n on the boundary of K_n such that $\Psi_{m0}(G_n) \cap (G_n + T_{K_n}(G_n)) = \emptyset$. In particular, $\Psi_{m0}(G_n) - G_n \subset [T_{K_n}(G_n)]^C$, where the superscript C denotes the complement of a set, and without loss of generality we will take $\Psi_{m0}(G_n)$ to be single-valued.

Now, let $\tilde{\Psi}_{0n}(G)$ denote the least-squares projection of $\Psi_{m0}(G)$ onto $K_n^{h_n} := \{G : d(G, K_n) \leq h_n\}$, the h_n-expansion of K_n. Note that if $\Psi_{m0}(G) - G \subset [T_{K_n}(G)]^C$, then also $\tilde{\Psi}_{0n}(G) - G \subset [T_{K_n}(G)]^C$. Now consider the sequence of vectors $v_n := h_n^{-1}\left(G_n - G^*_{0,j}, \tilde{\Psi}_{0n}(G_n) - G_n\right)$ in $K_n \times K_n^{h_n}$. Since this sequence is contained in the closed ball of radius 2 - a compact set - there exists a converging subsequence $v_{n(\nu)}$, $\nu = 1, 2, \ldots$ and $n(\nu) \to \infty$, where $\lim_{\nu\to\infty} v_{n(\nu)} =: v \equiv (v_G, v_\Psi)$. Since $0 \in \tilde{\Psi}_{0n}(G^*_{0,j}) - G^*_{0,j}$ by assumption, v is also an element of the graph of the contingent derivative of $\Psi_{m0}(G) - G$ at (G^*, G^*).

Furthermore, K_n is convex, so that $G^*_{0,j} - G_n \in T_{K_n}(G_n)$. Since K_n is a closed ball, at every point G_n on the boundary of K_n, $T_{K_n}(G_n)$ is a closed half-space. This implies that $[T_{K_n}(G_n)]^C = -\text{int}(T_{K_n}(G_n))$, where $\text{int}(A)$ denotes the interior of a set A. Now, by construction $G_n - \tilde{\Psi}_{0n}(G_n) \in [T_{K_n}(G_n)]^C$, so that the inner product $(G_n - G^*_{0,j})'(\tilde{\Psi}_{0n}(G_n) - G_n) > 0$ for all n. Hence, taking limits along the subsequence $n(\nu)$, $\nu = 1, 2, \ldots$, we have $v'_G v_\Psi \geq 0$. However, this contradicts that for the contingent derivative of Ψ_{m0} at $(G^*_{0,j}, G^*_{0,j})$ we had $u'v < 0$ whenever $v \in D\Gamma_0(G^*_{0,j}, G^*_{0,j})(u)$. Hence every h_n-neighborhood of $G^*_{0,j}$ must contain a compact convex subset $\bar{K}_j$ containing an open set around $G^*_{0,k}$ such that Ψ_{m0} has to be inward on $\bar{K}_j$ as claimed before.

Also, since $\bar{K}_j$ is compact, and $\Psi_{m0}(G) - G$ is continuous, we have that the length of the projection of $\Psi_{m0}(G) - G$ onto $T_{\bar{K}_j}(G)$ is bounded away from zero as G varies over $\bar{K}_j$. Noting that the convex hull of $\hat{\Psi}_{mn}(G)$ is lower semi-continuous and converges to $\Psi_{m0}(G)$ graphically in probability, we have that $\hat{\Psi}_{mn}$ is also inward on $\bar{K}_j$ with probability approaching one.

Local Existence of Fixed Points of $\hat{\Psi}_{mn}$. In order to remove the non-convexities of the mapping $\hat{\Psi}_{mn}(G)$, we consider a transformation of the graph of $\hat{\Psi}_{mn}$ on the set $\bar{K}_j$ under a continuous one-to-one mapping $H_j : \bar{K}_j \times \bar{K}_j \to G \times G$, $[G, \Psi] \mapsto [H_j(G, \Psi - G), \Psi]$. We can choose the mapping H_j such that $H_j(G, \Psi - G)$ is strictly monotone in its second argument, and $H_j(G, 0) = G$, and the values of $(H_j(G, \hat{\Psi}_{mn}(G)), \hat{\Psi}_{mn}(G))$ are convex for every $G \in \Delta\mathcal{S}$. Note that the two conditions on $H_j(\cdot)$ imply that G is a fixed point of $\hat{\Psi}_{mn}$ if and only if it is also a fixed point of $(H_j(G, \hat{\Psi}_{mn} - G), (\hat{\Psi}_{mn}(G))$. Furthermore, since the transformations H_j map any point on the boundary of $\bar{K}_j$ onto itself, they also preserve the inward mapping property of $\hat{\Psi}_{mn}$.

Since the transformed mapping is nonempty with closed and convex values, upper semi-continuous, and inward on $\bar{K}_j$, existence of a fixed point in $\bar{K}_j$ follows from Theorem 3.2.5 in Aubin and Frankowska (1990). Since the number of sets $K_1, \ldots, K_s$ is finite, and the set of fixed points of the transformed mapping is equal to that of $\hat{\Psi}_{mn}$, it follows that with probability approaching one, $\hat{\Psi}_{mn}$ has a fixed point in the neighborhood of each equilibrium point $G^*_{0,j} \in \mathcal{G}^*_{m0}$, establishing the conclusion in part (b) $\square$

C.1. **Proof of Lemma 3.2.** We first consider case (a). Consider a pair of actions $(s^{(k)}, s^{(l)})$, and let $\tilde{t} := (\tilde{t}_1, \tilde{t}_2) \in \mathcal{T}_1 \times \mathcal{T}_2$ such that a player of type $\tilde{t}$ is indifferent between these two or more actions given the aggregate state G. Note that t_{m1} is a finite-dimensional random variable in a Euclidean space and therefore its conditional distribution given t_{m2} is tight. Hence, for a fixed $\delta > 0$, we can find a compact subset $K \subset \mathcal{T}_1$ such that $t_{mi1} \in K$ with conditional probability greater than $1-\delta$ given $t_{mi2} = \tilde{t}_2$.

Since by Assumption 3.3 (ii), the Jacobian $\nabla_{t_{m1}}\mathbf{u}(G, t_{mi})$ is continuous and has rank equal to $p-1$, there exist a neighborhood $N(\tilde{t}_1)$ of $\tilde{t}_1$ and a vector $a(\tilde{t}_1) \in \mathbb{R}^{p-1} \setminus \{0\}$ such that the directional derivative $a(\tilde{t}_1)'\nabla_{t_{m1}}\left[u(s^{(k)}, G, t) - u(s^{(l)}, G, t)\right] > 0$ for all $t = (t_{m1}, \tilde{t}_2)$ such that $t_{m1} \in N(\tilde{t}_1)$. Consider the projection $B(\tilde{t_{m1}})t_{m1} := [I_{p-1} - a(\tilde{t}_1)(a(\tilde{t}_1)'a(\tilde{t}_1))^{-1}a(\tilde{t}_1)']t_{m1}$. By construction of $a(\tilde{t}_1)$ and $B(\tilde{t}_1)$, we have $B(\tilde{t}_1)a(\tilde{t}_1) = 0$ and that for every $b \in \mathbb{R}^{p-1}$ there exists at most one value of $t_{m1} \in N(\tilde{t}_1)$ such that $B(\tilde{t}_1)t_{m1} = b$ and $u(s^{(k)}, G, t) = u(s^{(l)}, G, t)$, where $t = (t_{m1}, \tilde{t}_2)$.

Therefore, since conditional on $\tilde{t}_2$, t_{m1} is continuously distributed with full support on $\mathbb{R}^{p-1}$, and $\mathcal{B} := \{B(\tilde{t}_1)t_{m1} : t_{m1} \in \mathbb{R}^{p-1}\}$ constitutes a proper linear subspace of $\mathbb{R}^{p-1}$, we have that $P(u(s^{(k)}, G, t) = u(s^{(l)}, G, t), t_{m1} \in N(\tilde{t}_1)|\tilde{t}_2, B(\tilde{t}_1)t_{m1}) = 0$ for all $b \in \mathcal{B}$. Hence by the law of total probability we can integrate out $B(\tilde{t}_1)t_{m1}$, so that the probability $P(u(s^{(k)}, G, t) = u(s^{(l)}, G, t), t_{m1} \in N(\tilde{t}_1)|\tilde{t}_2) = 0$.

Without loss of generality we can take the family of neighborhoods $N(\tilde{t}_1)$ to be a cover of K. Since K is compact, there is a finite subcover $N_1, \dots, N_J$, and by the law of total probability

$$P\left(u(s^{(k)}, G, t) = u(s^{(l)}, G, t), t_{m1} \in K \middle| \tilde{t}_2 \right) = \sum_{j=1}^{J} P\left(u(s^{(k)}, G, t) = u(s^{(l)}, G, t), t_{m1} \in N_j \middle| \tilde{t}_2 \right) = 0$$

for any given pair of action $s^{(k)}, s^{(l)}$. Since the number of pairs of actions $\binom{p}{2}$ is finite, we also have

$$\begin{aligned} P & \left(u(s^{(k)}, G, t) = u(s^{(l)}, G, t) \text{ for some } k \neq l, t_{m1} \in K \middle| \tilde{t}_2 \right) \\ & \leq \textstyle\sum_{k<l} P\left(u(s^{(k)}, G, t) = u(s^{(l)}, G, t), t_{m1} \in K \middle| \tilde{t}_2 \right) = 0 \end{aligned} \tag{C.1}$$

Since $\delta \geq P(t_{m1} \notin K|\tilde{t}_2)$ can be chosen arbitrarily small, we have by the law of total probability that $\psi_0(t_{mi}; G)$ is a singleton with probability 1, and the correspondence $\Psi_{m0}(G)$ is in fact single-valued.

For case (b), it is sufficient to notice that due to strict monotonicity of the payoff differences $u(s^{(l)}, G, t) - u(s^{(k)}, G, t)$, for every value of $\tilde{t}_2$ there is at most one value of t_{m1} such that a player of type $(t_{m1}, \tilde{t}_2')'$ is indifferent between actions $s^{(k)}$ and $s^{(l)}$. Since the number of actions is finite, the conditional probability of a tie given $t_2 = \tilde{t}_2$ is zero, and the conclusion follows from the law of total probability $\square$

C.2. **Proof of Proposition 3.2.** Let $\nu_{m1}, \nu_{m2}, \dots$ be a sequence of i.i.d. draws from the uniform distribution on the p-variate probability simplex $\Delta\mathcal{S} := \left\{\pi \in \mathbb{R}_+^p : \sum_{q=1}^p \pi_q = 1\right\}$, where p is the number of pure actions available to each player. Then $(\nu_{m1}', \dots, \nu_{mn}')'$ is uniformly distributed on

$$(\Delta\mathcal{S})^n = \left\{\pi \in \mathbb{R}_+^{n(p-1)} : \sum_{q=1}^{p} \pi_{iq} = 1 \text{ for all } i = 1, \dots, n\right\}.$$

Now fix $w_m \in \mathcal{W}_m$ and let $\lambda_{n_0}^*(w_m)$ be the distribution over Bayes Nash equilibria σ^* that generates the mixture $f_{n_0}^*(s, x|\theta)$ over $\Sigma^*(w_m)$. Without loss of generality we can assume that the ith coordinate of λ_{0n}^* is of the form $\lambda_{n_0,i}^*(w_m) := \tilde{\lambda}_{n_0}(\nu_{mi}, w_m)$, where the function $\tilde{\lambda}_{n_0}(\cdot, w_m)$ is symmetric with respect to permutations among players for whom the individual payoff-relevant information in w_m is the same. Then $\lambda_{n_0}^*$ constitutes a partition of $(\Delta\mathcal{S})^{n_0}$ that can be chosen to be symmetric with respect to the same permutations of indices $1, \dots, n_0$, and therefore satisfies the symmetry assumptions on the equilibrium selection rule $\lambda_n(\cdot)$. In particular, the resulting distribution of type-action characters, $f(s, x|\theta, \lambda_{n_0}^*) = f_{n_0}^*(s, x|\theta)$ for all $s \in \mathcal{S}$.[30]

[30]We can in general map every Bayes Nash equilibrium to $\bar{\sigma}(\sigma^*) \in (\Delta\mathcal{S})^n$ where the ith coordinate is given by $\bar{\sigma}_i(\sigma^*) := \mathbb{E}[\sigma_{mi}^*(t_{mi}, w_m)|w_m]$. We can then find a partition (up to shared boundaries) $(\Delta\mathcal{S})^n = \bigcup_{\sigma^*} V_n(\sigma^*)$ such that for any BNE σ^*, $\bar{\sigma}(\sigma^*) \in V_n(\sigma^*)$, and $P((\nu_{m1}, \dots, \nu_{mn}) \in V_n(\sigma^*)) = \lambda_{n_0}^*(w_m)$. Note that if for two players i, j the public signal w_m contains the same player-specific information about types, $H_m(t_{mi}|w_m) =$

We now prove part (ii). By Theorem 3.1, we have that the equilibrium value of the aggregate $d(G^*_{mn}, \mathcal{G}^*_{m0}) \stackrel{a.s.}{\to} 0$, where $\mathcal{G}^*_{m0}$ is a finite set. Suppose we now fix n_1 and w_m, and let $\tilde{G}^* := \arg\min_{G \in \mathcal{G}^*_{m0}} d(G, G^*_{n_1})$, where $G^*_{n_1}$ is the value of the equilibrium aggregate under the selection rule λ_{n_1}. By Theorem 3.1 (b), for each $\eta > 0$ there exists n_1 large enough (but not necessarily larger than n_0) such that with probability $1 - \eta$ we can find an equilibrium that supports some $\tilde{G}^*_n$ for all $n \geq n_1$ such that $d(\tilde{G}^*_n, \tilde{G}^*) < r_n$ a.s. for a sequence $r_n \to 0$. Hence we can choose the sequence $\lambda_n(\cdot)$ such that for the first n_1 coordinates of $\nu_{m1}, \dots, \nu_{mn}$ and $t_{m1}, \dots, t_{mn}$ coinciding with $\nu_{m1}, \dots, \nu_{n_1}$ and $t_{m1}, \dots, t_{n_1}$, respectively, we have $G^*_n = \tilde{G}^*_n$ for all $n \geq n_1$.

Now let $M := \sup_{\theta \in \Theta} \max_{s,s' \in \mathcal{S}} |\mathbb{E}[m((s,x);\theta) - m((s',x);\theta)]|$, which is finite by assumption. Now we can bound

$$\mathbb{E}\left[|m(y_{m1,n};\theta) - m(y_{m1};\theta)| \,|\, \mathcal{F}_n\right] \leq MP\left(s_{n1} \neq s_1 | \mathcal{F}_n\right)$$

Now notice that by Assumption 3.3 and Lemma 3.2, the set $\psi_0(t_{mi}; \tilde{G}^*)$ is the singleton $\{s_{mi}\}$ with probability one. Therefore since $t_{m1}, \dots, t_{mn}$ are i.i.d. draws from $H_0(t)$, we have

$$P(s_{n1} \neq s_1 | \mathcal{F}_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left\{ \psi_0(t_{mi}; \tilde{G}^*) \neq \psi_0(t_{mi}; G^*_{mn}) \right\} \leq 2P(\psi_0(t_{m1}; \tilde{G}^*) \neq \psi_0(t_{m1}; G^*_{mn}))$$

say, w.p.a.1 by the strong law of large numbers. Now notice that by Assumption 3.3, $P(\psi_0(t_{mi}; \tilde{G}^*) \neq \psi_0(t_{mi}; G))$ is continuous in G so that, since $G^*_{mn} \to \tilde{G}^*$ a.s., we can make $P(\psi_0(t_{mi}; \tilde{G}^*) \neq \psi_0(t_{mi}; G^*_{mn}))$ arbitrarily small by choosing n large enough, establishing requirement (ii), where n_1 and c_n can be chosen independently of $\hat{G}^*_{n_0}$ and $f^*_{n_0}(s, x|\theta)$ $\square$

Appendix D. Proofs for Results from Section 5

D.1. **Proof of Theorem 5.1.** Since the assumptions of this theorem subsume the conditions for Proposition 3.2, we can find a sequence of equilibrium selection rules $\{\lambda_n\}_{n \geq 1}$, $\lambda_n \in \Lambda_n(\theta)$ that is of the form $\lambda_{ni} = \lambda_n(\nu_{mi}, w_m)$. Since the augmented types $(x_{mi}, \varepsilon_{mi}, \nu_{mi})_{i \geq 1}$ are exchangeable and λ_n is invariant to permutations of the agent-specific information in w_m, the resulting observable type-action characters $y_{mi,n} := (s_{mi,n}, x'_{mi})'$ are also exchangeable for every n.

Furthermore, it follows from Proposition 3.2 that

$$\sup_{G \in \Delta \mathcal{S}} \left|\mathbb{E}\left[m(y_{m1,n}, G; \theta) - m(y_{m1}, G; \theta) \,|\mathcal{F}_n\right]\right| < c_n \quad \text{a.s.}$$

for all n and $\theta \in \Theta$, and a deterministic null sequence $c_n \to 0$ that does not depend on the cross-sectional distribution at n_0. Furthermore, $G^*_{mn} \stackrel{a.s.}{\to} G^*_{m0}$ by Theorem 3.1, and $\hat{G}_{mn} - G^*_{mn} \stackrel{a.s.}{\to} 0$ by the strong law of large numbers since types are conditionally i.i.d. given $\mathcal{F}_\infty$. Since $\mathbb{E}\left[m(y_{m1}, G; \theta) | \mathcal{F}_\infty\right]$ is also continuous in G and the number of equilibrium points G^*_{m0} is finite by Lemma 3.1, we also have that

$$\left|\mathbb{E}\left[m(y_{m1,n}, \hat{G}_{mn}; \theta) - m(y_{m1}, G^*_{m0}; \theta) \,|\mathcal{F}_n\right]\right| \to 0 \text{a.s.} \tag{D.1}$$

Also, by Proposition 3.2, the column-wise limit of $y_{mi,n}$ as $n \to \infty$ is well-defined with probability one for every $i = 1, 2, \dots$, so that n-exchangeability of $y_{m1,n}, \dots, y_{mn,n}$ implies that the limiting sequence $(y_{mi})_{i \geq 1}$ is infinitely exchangeable.

$H_m(t_j|w_m)$, then the coordinate sets $\{\bar{\sigma}_i(\sigma^*)\}$ and $\{\bar{\sigma}_j(\sigma^*)\}$ are the same, so that w.l.o.g. we can choose the sets $V_n(\sigma^*)$ to be symmetric in its ith and jth coordinate. Noting that $\nu_{m1}, \nu_{m2}, \dots$ are w_m-measurable by assumption, we can choose the ith component of $\lambda_n(w_m)$ as $\lambda_{ni}(w_m) = \tilde{\lambda}_n(\nu_{mi}, w_m) := \sigma^*$ if and only if $(\nu_{m1}, \dots, \nu_{mn}) \in V_n(\sigma^*)$, so that the implied distributions for $n = n_0$ are the same under $\lambda^*_{n_0}(w_m)$ and $\lambda_{n_0}(\cdot, w_m)$.

Now a standard argument (see e.g. Kingman (1978)) yields that for any n-symmetric event $A_n \in \mathcal{F}_n$ and any $j = 1, \dots, n$,

$$\mathbb{E}[m(y_{m1}, G^*_{m0}; \theta)\mathbb{1}_{A_n}] = \mathbb{E}[m(y_j, G^*_{m0}; \theta)\mathbb{1}_{A_n}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n m(y_{mi}, G^*_{m0}; \theta)\mathbb{1}_{A_n}\right]$$

Since $\frac{1}{n}\sum_{i=1}^n m(y_{mi}, G^*_{m0}; \theta)\mathbb{1}_{A_n}$ is $\mathcal{F}_n$-measurable, we obtain

$$\mathbb{E}[m(y_{m1}, G^*_{m0}; \theta)|\mathcal{F}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n m(y_{mi}, G^*_{m0}; \theta)\right| \mathcal{F}_n\Bigg] = \frac{1}{n}\sum_{i=1}^n m(y_{mi}, G^*_{m0}; \theta)$$

Hence by Condition 5.1, the sequence

$$Z^*_n := \frac{1}{n}\sum_{i=1}^n m(y_{mi,n}, \hat{G}_{mn}; \theta) + \frac{1}{n}\sum_{i=1}^n \left[m(y_{mi}, G^*_{m0}; \theta) - m(y_{mi,n}, \hat{G}_{mn}; \theta)\right]$$

satisfies a uniform integrability condition. Since in addition

$$\begin{aligned}\mathbb{E}[Z^*_n|\mathcal{F}_{n+1}] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n m(y_{mi}, G^*_{m0}; \theta)\right| \mathcal{F}_{n+1}\Bigg] = \mathbb{E}\left[m(y_{m1}, G^*_{m0}; \theta)|\, \mathcal{F}_{n+1}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n+1} m(y_{mi}, G^*_{m0}; \theta) = Z^*_{n+1},\end{aligned}$$

Z^*_n is a reverse martingale adapted to the filtration $\{\mathcal{F}_n\}_{n=1}^\infty$, so that by the reverse martingale theorem (e.g. Theorem 2.6 in Hall and Heyde (1980)), $Z^*_n \stackrel{a.s.}{\to} \mathbb{E}[m(y_{m1}, G^*_{m0}; \theta)|\mathcal{F}_\infty]$.

On the other hand, (D.1) implies that

$$\left|Z^*_n - \frac{1}{n}\sum_{i=1}^n m(y_{mi,n}, \hat{G}_{mn}; \theta)\right| = \left|\mathbb{E}\left[m(y_{mi}, G^*_{m0}; \theta) - m(y_{mi,n}, \hat{G}_{mn}; \theta)|\mathcal{F}_n\right]\right| \to 0\text{a.s.}$$

so that

$$\frac{1}{n}\sum_{i=1}^n m(y_{mi,n}, \hat{G}_{mn}; \theta) \stackrel{a.s.}{\to} \mathbb{E}[m(y_{m1}, G^*_{m0}; \theta)|\mathcal{F}_\infty]$$

Furthermore, almost sure convergence is joint for any random vector satisfying the conditions of this Theorem, so that uniformity in θ follows from Condition 5.1 following the same reasoning as in the proof of the (almost sure) Glivenko-Cantelli theorem for the i.i.d. case, see van der Vaart (1998), Theorems 19.1 and 19.4 □

D.2. **Proof of Theorem 5.2.** To establish the conclusion of this theorem, we verify the conditions for Theorem 3.3.1 in van der Vaart and Wellner (1996) to derive the asymptotic distribution of $\sqrt{n}(\hat{\mu} - \mu_0)$. Note that for notational simplicity, we suppress the θ argument below whenever θ can be regarded as fixed. For a vector b and a set A, we also denote $A + b := A \oplus \{b\}$ to simplify notation.

First, note that under Assumption 3.3, $\Psi_{m0}(G)$ is single-valued at every point $G \in \Delta\mathcal{S}$ by Lemma 3.2. Hence, the Aumann selection expectation defining $\Psi_{m0}(G)$ coincides with the usual expectation of a vector-valued random variable, and under Assumption 3.3 (i) the mapping $G \mapsto G - \Psi_{m0}(G)$ is Fréchet differentiable with continuous derivative $I_{p-1} - \dot{\Psi}_G$.

By assumption 5.2 (iv), the Jacobian

$$\dot{M}_G(G) := \nabla_G \mathbb{E}[m(y_{mi}, G; \theta)|\mathcal{F}_\infty] = \mathbb{E}[\nabla_G m(y_{mi}, G; \theta)|\mathcal{F}_\infty]$$

is well-defined and bounded. Similarly, by Assumption 3.3 (iii), the derivative of the conditional probability

$$\dot{\Psi}_G(s, G; x_{mi}) := \nabla_G P\left(\delta_s \in \psi_0(t_{mi}; G) \,|\, x_{mi}, \mathcal{F}_\infty\right)$$

exists and is bounded and continuous in G a.s.. Hence we can interchange differentiation and integration, so that by the law of iterated expectations

$$\dot{B}_G(G) := \nabla_G \mathbb{E}\left[m((\psi_0(t_{mi}; G)', x_{mi}')', G; \theta) \,|\, \mathcal{F}_\infty\right] := \dot{M}_G(G) + \sum_{s \in \mathcal{S}} \mathbb{E}\left[m((s, x_{mi}')'; \theta)\dot{\Psi}_G(s, G; x_{mi}) | \mathcal{F}_\infty\right]$$

Since $\dot{\Psi}_G(G; x_{mi})$ is a.s. bounded, and by Assumption 5.2 (ii), the functions $m((s, x_{mi}')', G; \theta)$ are dominated by a square integrable envelope for all $s \in \mathcal{S}$, we can use the Cauchy-Schwarz inequality to establish that the elements of $\nabla_G \mathbb{E}\left[m((\psi_0(t_{mi}; G)', x_{mi}')', G; \theta) \,|\, \mathcal{F}_\infty\right]$ are bounded and continuous in G in a neighborhood of G_{m0}^*.

Next consider the difference between the maps $\hat{r}_n(\mu) := \hat{r}_n(\theta, \mu)$. Let $r_n^*(t_{mi}; \mu)$ be an arbitrary selection of the multi-valued function

$$r_n^*(t_{mi}; \mu) \in \begin{bmatrix} m((\psi_0(t_{mi}', G), x_i')', \hat{G}_{mn}(G); \theta) - m \\ \mathbb{E}\left[\psi_0(t_{mi}; G) \,|\, w_m\right] - G \end{bmatrix}$$

We can now decompose the difference between the maps $\hat{r}_n(\mu) := \hat{r}_n(\theta, \mu)$ and $r_0(\mu) := r_0(\theta, \mu)$ into

$$\begin{aligned} \sqrt{n}(\hat{r}_n(\mu_0) - r_0(\mu_0)) \quad &= \quad \frac{1}{\sqrt{n}} \bigoplus_{i=1}^n \left[r_n(t_{mi}; \mu_0) - r_n^*(t_{mi}; \mu_0)\right] + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[r_n^*(t_{mi}; \mu_0) - r(t_{mi}; \mu_0)\right] \\ &\quad\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[r(t_{mi}; \mu_0) - \mathbb{E}\left[r(t_{mi}; \mu_0) | \mathcal{F}_\infty\right]\right] \\ &=: \quad T_1 + T_2 + T_3 \end{aligned}$$

We first show that the first term is asymptotically negligible: Define $M := \sup_{\theta,G} \max_{s,s' \in \mathcal{S}} |\mathbb{E}[m((s,X),G;\theta) - m((s',X),G;$ which is finite by Assumption 5.1. If we let $L_n := \{G' \in \Delta\mathcal{S} : \|G_0 - G'\| \leq \frac{1}{n}\}$ and $\pi_n(\mu_0) := P\big(\psi_0(t_{mi}; G_0) \neq \bigcup_{G' \in L_n} \psi_0^*(t_{mi}$ then we can bound

$$\left\|\mathbb{E}\left[\left|(r_n^*(t_{mi}; \mu_0) - r(t_{mi}; \mu_0))^2\right| \mathcal{F}_\infty\right]\right\| \leq \pi_n(\mu_0)(1 + M)$$

Similarly, we can bound

$$|\mathbb{E}[r_n^*(t_{mi}; \mu_0) - r(t_{mi}; \mu_0) | \mathcal{F}_\infty]| \leq \pi_n(\mu_0)$$

Now notice that, as shown above Assumption 3.3 implies that at every value of G, there are no atoms of "switchers" in the type distribution. Since the density of the type distribution is bounded from above, we therefore have that $\lim_n n\pi_n(\mu_0) < \infty$ for each μ_0. Since with probability one, μ_0 only takes finitely many values, the common upper bound for all μ_0 is finite for all $\mathcal{F}_\infty$-measurable events. Therefore the variance and absolute value of the expectation of the first term $T_1 := \frac{1}{\sqrt{n}} \bigoplus_{i=1}^n \left[r_n(t_{mi}; \mu_0) - r_n^*(t_{mi}; \mu_0)\right]$ converge to zero almost surely as $n \to \infty$, so that applying Markov's inequality, T_1 converges to zero in probability conditional on $\mathcal{F}_\infty$.

For the second term, notice that the components in $r_n^*(t; \mu)$ corresponding to the equilibrium conditions are equal to those in $r(t; \mu)$ so that we only have to account for the components corresponding to the moment functions. We can use a mean-value expansion of the moment function $m(y, G; \theta)$ with respect to G to show that

$$\begin{aligned} T_2 &:= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[m((\psi_0(t'_{mi},G),x'_i)',\hat{G}_{mn}(G);\theta) - m((\psi_0(t'_{mi},G),x'_i)',G;\theta)\right] \\ &= \left[\frac{1}{n}\sum_{i=1}^n \nabla_G m((\psi_0(t'_{mi},G),x'_i)',G;\theta)\right]\sqrt{n}(\hat{G}_{mn}(G)-G)+o_p(1) \end{aligned}$$

By the conditional law of large numbers in Theorem 5.1, $\frac{1}{n}\sum_{i=1}^n \nabla_G m((\psi_0(t'_{mi},G),x'_i)',G;\theta) \to \dot{M}_G(G)$ almost surely, which is continuous in G by assumption, and furthermore

$$\sqrt{n}(\hat{G}_{mn}(G)-G) = \frac{1}{\sqrt{n}}\left(\sum_{i=1}^n \psi_0(t_i;G) - G\right) + o_p(1)$$

by an analogous argument to our analysis of the term T_1.

Since $\psi_0(t_{mi};G^*_{m0})$ is single-valued with probability 1, and t_{mi} are i.i.d. draws from the distribution $H_m(t|w_m)$, we can apply a standard CLT and obtain

$$T_2+T_3 := \frac{1}{\sqrt{n}}\sum_{i=1}^n [r^*_n(t_{mi};\mu_0) - r_0(\mu_0)] \stackrel{d}{\to} N(0,\Omega(G^*_{m0})) \tag{D.2}$$

Since conditional on $\mathcal{F}_\infty$, $T_1 \to 0$ in probability, convergence in distribution with mixing follows from Slutsky's theorem.

Next, consider the Jacobian of the (augmented) population moments,

$$\nabla_\mu r_0(\mu) = -\left[\begin{array}{cc} I_q & -\dot{B}_0(G^*_{m0}) \\ 0 & I_{p-1} - \dot{\Psi}_{m0}(G^*_{m0}) \end{array}\right]$$

By Assumption 3.2, the derivative $I_{p-1} - \dot{\Psi}_{m0}(G^*_{m0})$ has full rank at every $G^*_{m0} \in \mathcal{G}^*_{m0}$, and we already showed above that $\dot{B}_0(G)$ and $\dot{\Psi}_{m0}(G)$ are continuous in G. Due to the block-diagonal structure it is therefore straightforward to verify that the inverse of $\nabla_\mu r_0(\mu)$ is well-defined and continuous in μ in a neighborhood around μ_0 and for all values of $\theta \in \Theta$.

We can now verify the conditions of Lemma 3.3.5 in van der Vaart and Wellner (1996). By Assumption 3.1 (iii), the class $\mathcal{H} := \{\psi_0(t_{mi};G) : G \in \Delta\mathcal{S}\}$ is a VC subgraph class, and therefore Donsker. Also, by Assumption 5.2, $\mathcal{M}_s := \{m((s,x),G;\theta), G \in \Delta\mathcal{S}, \theta \in \Theta\}$ is a Donsker class for each $s \in \mathcal{S}$, and $\mathcal{S}$ is finite. Note that the class $\mathcal{R} := \{r(t_{mi};\mu,\theta), \mu \in \mathbb{R}^q \times \Delta\mathcal{S}, \theta \in \Theta\}$ is a Lipschitz transformation of $\mathcal{M}_1,\dots,\mathcal{M}_p$ and $\mathcal{H}$ with the same envelope function as $\mathcal{M}_1$. Hence, by Theorem 2.10.6 in van der Vaart and Wellner (1996), $\mathcal{R}$ is also Donsker with respect to the distribution of types.

Note also that as shown before, $\psi_0(t_{mi};G,s)$ is single-valued at every G except for a set of types of measure zero. Since payoffs are continuous in t and G, and the p.d.f. of the type distribution is continuous in t, it follows from Assumption 3.3 and Lemma 3.2 that $\mathbb{E}\|r(t_{mi};\mu) - r(t_{mi};\mu_0)\|^2$ is also continuous in μ.

Finally, by Proposition 3.2, we can construct the coupling such that conditional on $\mathcal{F}_\infty$, $\hat{G}_{mn} - G_0 \to 0$ in outer probability, so that by Lemma 3.3.5 in van der Vaart and Wellner (1996) we have that for the empirical process

$$\|\mathbb{G}_n(r(\cdot,\mu) - r(\cdot,\mu_0))\| = o^*_P\left(1+\sqrt{n}\|\mu-\mu_0\|\right)$$

Hence we can apply Theorem 3.3.1 in van der Vaart and Wellner (1996) to obtain the asymptotic distribution of $\sqrt{n}(\hat{\mu}_n - \mu_0)$.

It remains to show uniformity with respect to θ. Note that since $\mathcal{R}$ is Donsker as θ varies in Θ, convergence of $\hat{r}_n(\theta,\mu)$ is also uniform in θ. Finally, since by Assumption 5.2 (ii), the class $\mathcal{M}_s$ has a square-integrable envelope function that does not depend on θ, by the same arguments as before the elements of the derivative $\dot{B}_G$ are bounded so that the bounds in the previous arguments do not depend on θ $\square$

D.3. **Proof of Corollary 5.2.** By the same reasoning as in the proof of Theorem 5.2, the matrices $\dot{B}_G$, $\dot{M}_G$, and $(I - \dot{\Psi}_G)^{-1}$ are well-defined, so that by Assumption 5.3 (i), the first two moments of $|v_{mi}|$ are bounded. Also, $\hat{v}_{mi,n}$ is continuous in $\hat{m}_n$, G^*_{mn}, $\widehat{\dot{B}}_G$, $\widehat{\dot{M}}_G$, and $\widehat{\dot{\Psi}}_G$ which converge to their limits m_0, G^*_{m0}, $\dot{B}_G$, $\dot{M}_G$, and $\dot{\Psi}_G$ respectively. Therefore, we can apply Theorem 5.1 to establish that $\frac{1}{n}\sum_{i=1}^n \hat{v}_{mi,n}\hat{v}'_{mi,n} \to \mathbb{E}[v_{mi}v'_{mi}|\mathcal{F}_\infty]$ a.s.. Next, the covariance matrix of $(m(y_{mi,n}, \hat{G}_{mn}; \theta)', \hat{\psi}'_{in})'$ is given by $\Omega_n(G^*_{mn})$ where $\Omega_n(G) := \mathrm{Var}(m((\psi^*_n(t_{mi}; G), x'_{mi})', \hat{G}_{mn}(G); \theta), \mathbb{E}[\psi^*_n(t_{mi}; G)|w_m]|\mathcal{F}_\infty)$ for some selection $\psi^*_n(t_{mi}; G) \in \psi_n(t_{mi}; G)$. By similar arguments as in the proof of Theorem 5.2, $\Omega_n(G) \to \Omega(G)$ uniformly in G. Note that $\Omega(G)$ is continuous in G and $G^*_{mn} \to G^*_{m0}$ a.s., so that $\Omega(G^*_{mn}) \to \Omega$ by the continuous mapping theorem.

Since by Assumption 5.3, the estimated Jacobians converge to $\dot{B}_G$, $\dot{M}_G$, and $\dot{\Psi}_G$, respectively, we therefore have that the conditional expectation of the product $v_{mi}v'_{mi}$ converges to $A'\Omega A$. By Assumption 5.3 (i) we therefore obtain $\hat{V}_n \to A'\Omega A$ almost surely, as claimed in the first claim of the Corollary. The second part follows immediately from Theorem 5.2 and Slutsky's Lemma $\square$

References

Andrews, D. (2005): "Cross-Section Regression with Common Shocks," *Econometrica*, 73(5), 1551–1585.

Aubin, J., and H. Frankowska (1990): *Set-Valued Analysis*. Birkhäuser.

Bajari, P., J. Hahn, H. Hong, and G. Ridder (2011): "A Note on Semiparametric Estimation of Finite Mixtures of Discrete Choice Models with Application to Game Theoretic Models," *International Economic Review*, 52(3), 807–824.

Bajari, P., H. Hong, and S. Ryan (2010): "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica*, 78(5), 1529–1568.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79(6), 1785–1821.

Berry, S. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60(4), 889–917.

Bickel, P., C. Klaassen, Y. Ritov, and J. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Springer.

Blum, J., H. Chernoff, M. Rosenblatt, and H. Teicher (1958): "Central Limit Theorems for Interchangeable Processes," *Canadian Journal of Mathematics*, 10, 222–229.

Blume, L. (1993): "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, 5, 387–424.

Bresnahan, T., and P. Reiss (1990): "Entry in Monopoly Markets," *Review of Economic Studies*, 57(4).

——— (1991): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48, 57–81.

Brock, W., and S. Durlauf (2001): "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68.

Chamberlain, G. (2010): "Binary Response Models for Panel Data: Identification and Information," *Econometrica*, 78(1), 159–168.

Chen, X., E. Tamer, and A. Torgovitsky (2011): "Sensitivity Analysis in a Semi-parametric Likelihood Model: A Partial Identification Approach," working paper, Yale and Northwestern.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791–1828.

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

Efron, B. (1979): "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7(1), 1–26.

Galichon, A., and M. Henry (2011): "Set Identification in Models with Multiple Equilibria," *Review of Economic Studies*, 78, 1264–1298.

Glaeser, E., B. Sacerdote, and J. Scheinkman (1996): "Crime and Social Interactions," *Quarterly Journal of Economics*, 111, 507–548.

Graham, B. (2008): "Identifying social interactions through conditional variance restrictions," *Econometrica*, pp. 643–660.

Grieco, P. (2012): "Discrete Games with Flexible information Structures: An Application to Local Grocery Markets," working paper, Penn State.

Hahn, J., and W. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72(4), 1295–1319.

Hall, P., and C. Heyde (1980): *Martingale Limit Theory and its Application.* Academic Press, New York.

Harsanyi, J. (1973): "Oddness of the Number of Equilibrium Points: A New Proof," *International Journal of Game Theory*, 2(1), 235–250.

Hausman, J. (1983): *Specification and Estimation of Simultaneous Equation Models*chap. 7, pp. 391–448. in: Grilliches and Intrilligator (eds.): Handbook of Econometrics, Vol. I.

Horowitz, J., and W. Härdle (1996): "Direct Semiparametric Estiamtion of Single-Index Models with Discrete Covariates," *Journal of the American Statistical Association*, 91, No. 436, 1632–1640.

Kalai, E. (2004): "Large Robust Games," *Econometrica*, 72(6), 1631–1665.

Kallenberg, O. (2005): *Probabilistic Symmetries and Invariance Principles.* Springer.

Kellogg, R. (1976): "Uniqueness in the Schauder Fixed Point Theorem," *Proceedings of the American Mathematical Society*, 60, 207–240.

Kingman, J. (1978): "Uses of Exchangeability," *Annals of Probability*, 6(2), 183–197.

Lazzati, N. (2012): "Treatment Response with Social Interactions," working paper, University of Michigan.

Manski, C. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205–228.

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

Mas-Colell, A., M. Whinston, and J. Green (1995): *Microeconomic Theory*. Oxford University Press.

Matzkin, R. (1992): "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica*, 60(2), 239–270.

McLennan, A. (1997): "The Maximal Generic Number of Pure Nash Equilibria," *Journal of Economic Theory*, 72, 408–410.

Menzel, K. (2013): "Large Matching Markets as Two-Sided Demand Systems," working paper, New York University.

Milgrom, P., and J. Roberts (1990): "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 58(6), 1255–1277.

Milgrom, P., and R. Weber (1985): "Distributional Strategies for Games with Incomplete Information," *Mathematics of Operations Research*, 10(4), 619–632.

Molchanov, I. (2005): *Theory of Random Sets*. Springer, London.

Munshi, K., and J. Myaux (2006): "Social Norms and the Fertility Transition," *Journal of Development Economics*, 80, 1–38.

Nakajima, R. (2007): "Measuring Peer Effects on Youth Smoking Behaviour," *Review of Economic Studies*, 74, 897–935.

Newey, W., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," Handbook of Econometrics, Vol IV Chapter 36.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2006): "Moment Inequalities and their Application," working paper, Harvard University.

Powell, J., J. Stock, and T. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57(6), 1403–1430.

Quenouille, M. (1956): "Notes on Bias in Estimation," *Biometrika*, 43(3/4), 353–360.

Rilstone, P., V. Srivastava, and A. Ullah (1996): "The Second-Order Bias and Mean Squared Error of Nonlinear Estimators," *Journal of Econometrics*, 75, 369–395.

Rockafellar, R., and R. Wets (1998): *Variational Analysis*. Springer, Heidelberg.

Shang, Q., and L. Lee (2011): "Two-Step Estimation of Endogenous and Exogenous Group Effects," *Econometric Reviews*, 30(2), 173–207.

Soetevent, A., and P. Kooreman (2007): "A Discrete-Choice Model with Social Interactions: with an Application to High School Teen Behavior," *Journal of Applied Econometrics*, 22(3), 599–624.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147–165.

Todd, P., and K. Wolpin (2012): "Estimating a Coordination Game in the Classroom," working paper, UPenn.

Topa, G. (2001): "Social Interactions, Local Spillovers and Unemployment," *Review of Economic Studies*, 68, 261–295.

van der Vaart, A. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge.

van der Vaart, A., and J. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer, New York.

Weintraub, G., L. Benkard, and B. van Roy (2008): "Markov Perfect Industry Dynamics With Many Firms," *Econometrica*, 78(6), 1375–1711.