The Pennsylvania State University

The Graduate School

College of the Liberal Arts

COLLISION AND COLLUSION: TESTING THE NASH BEHAVIOR IN ENTRY GAMES OF COMPLETE INFORMATION

A Dissertation in

Economics

by

Nail Kashaev

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2016

The dissertation of Nail Kashaev was reviewed and approved* by the following:

Joris Pinkse
Professor of Economics
Dissertation Co-Advisor
Co-Chair of Committee

Sung Jae Jun
Associate Professor of Economics
Dissertation Co-Advisor
Co-Chair of Committee

Andrés Aradillas-López
Associate Professor of Economics

Edward Jaenicke
Associate Professor of Agricultural Economics

Barry Ickes
Professor of Economics
Head of the Department of Economics

*Signatures are on file in the Graduate School.

Abstract

The present dissertation consists of three essays on Econometrics of discrete outcome models. The first essay analyses possibility of discriminating between several solution concepts in a general class of semiparametric finite games with complete information based on observed data on outcomes and characteristics of agents. I find conditions under which it is possible to identify whether actual behavior of agents is consistent with a given solution concept. I propose different applications for my general methodology. For example, I can identify whether and how often firms play *Nash equilibria* (NE) in an entry game, which equilibria are more likely to be selected, and whether profit functions are private information or common knowledge. I also identify whether choices are sequential or simultaneous.

The second essay is a logical continuation of the first one. I focus on entry games with complete information and provide a statistical tool to test for the NE solution concept. I develop a sieve likelihood ratio type procedure to test whether the NE assumption can rationalize the data on outcomes and payoff shifters in semiparametric entry games with *second-order rational* firms. I allow agents to play mixed strategy NE in the regions of NE multiplicity. I do not impose parametric restrictions on the way firms randomize between different equilibria. The testing procedure does not assume that the model is point identified. I apply the proposed procedure to the data on entry and exit decisions of small grocery stores in rural areas in the USA.

In the third essay I address the issue of data availability in parametric binary outcome models. In particular, I consider environments where the researcher only observes the data that correspond to a particular outcome (pure choice-based data) and has some auxiliary information about the distribution of the explanatory variables. I propose a Generalized Method of Moments type procedure to estimate parametric binary models with pure choice-based data when auxiliary information on distribution of explanatory variables is available. As an empirical application of my procedure, I estimate the probability of a two-car collision based on the data on all police reported accidents in Seattle from 2002-2011.

Contents

List of Figures

Acknowledgements

I would like to express a deep gratitude to my advisers, teachers and mentors, Joris Pinkse and Sung Jae Jun. Their support and guidance have been an essential part of my work at Penn State.

I also would like to thank Andrés Aradillas-López, Paul Grieco, Patrik Guggenberger and Marc Henry for constant encouragement and numerous suggestions.

I am very grateful to to the rest of the faculty that has helped and taught me over the years including Herman Bierens, Jonathan Eaton, Ron Gallant, Edward Green, James Jordan, Edward Jaenicke, Kala Krishna, Vijay Krishna, Rohit Lamba, Mark Roberts, Spiro Stefanou, James Tybout, Quang Vuong, Neil Wallace, Stephen Yeaple, Ruilin Zhou, as well as conference and seminar presenters and attendants who have provided valuable comments and suggestions, specially Francesca Molinari and Marcelo Moreira.

My work would not be accomplished without my friend and coauthor Bruno Salcedo. He has been my third "advisor" for these six years at Penn State.

I am also grateful to other classmates and friends who have helped me and motivated me, including Bulat Gafarov, Gustavo Gudiño, Lidia Kosenkova, Huihui Li, Kai Li, Evelyn Nunes, Leonid Ogrel, Mark Razhev, Suraj Shekhar, Bruno Sultanum, Oleksandr Yaroshenko.

Lastly, I would like to thank my family, my brother Kamil, aunt Gulshat and uncle Rinat. My parents have been supporting my pursuit of academic career from the very beginning. I dedicate this dissertation to them.

All remaining errors are my own.

To my parents: Rishat Kashaev and Guzel Emasova

Chapter 1

Identifying solution concepts

Abstract This essay is based on joint work with Bruno Salcedo, Kashaev and Salcedo (2016). Empirical analyses of discrete games relies on behavioral assumptions that are crucial not just for estimation, but also for the validity of counterfactual exercises and policy implications. We find conditions for a general class of complete-information games under which it is possible to identify whether actual behavior satisfies some of these assumptions. We propose different applications for our general approach. For instance, our results allow us to identify whether and how often firms in an entry game play Nash equilibria, which equilibria are more likely to be selected, whether they use mixed strategies, whether they make choices simultaneously or sequentially, and whether profit functions are private information or common knowledge.

Existing methodologies to recover structural parameters from observed choices in discrete games rely on assumptions about the environment and the nature of strategic behavior, which can often be captured by the *solution concepts* being used.[1] We ask when it is possible to test whether actual choices satisfy such assumptions, and we find rather encouraging results. Under our conditions, a solution concept can rationalize the observed choices if and only if the actual behavior of the agents satisfies them

[1]Relevant examples include papers about discrete complete-information games (Bjorn and Vuong, 1984, Tamer, 2003, Aradillas-López and Tamer, 2008, Bajari et al., 2010b, Galichon and Henry, 2011, Henry and Mourifie, 2012, Kline and Tamer, 2012, Aradillas-López and Rosen, 2013), incomplete-information games (Bajari et al., 2010a, Aradillas-López, 2010, De Paula and Tang, 2012, Grieco, 2014), and entry games (Bresnahan and Reiss, 1990, 1991, Berry and Tamer, 2006, Ciliberto and Tamer, 2009, Kline, 2015a).

almost surely (Theorem 1.5). Moreover, competing solution concepts can be distinguished as long as they make disjoint predictions with positive probability (Theorem 1.6).

The problem of choosing a solution concept for each particular situation may sometimes be better approached from an empirical perspective, by analyzing choice patterns within the specific context.[2] Moreover, the importance of choosing the right solution concept goes beyond estimation. Even if structural parameters could be consistently estimated assuming invalid solution concepts, both the positive predictions and the normative implications of a model may depend not just on the characteristics of the environment, but also on behavioral or strategic considerations. Thus, our results are also relevant for establishing the validity of counterfactual analyses and policy implications.

Our framework can be applied to general discrete complete-information games either in extensive or strategic form. We illustrate our approach using a simple two-firm entry model adapted from Bresnahan and Reiss (1990). We show that it is possible to test whether firms play Nash equilibria (NE), and we provide a specific example in which incorrectly assuming Nash play leads to completely misleading policy recommendations, even if the parameters of the model are known. We also identify whether firms use mixed strategies, and whether choices are sequential or simultaneous. For our particular entry game, we are able to identify whether payoffs are private information or common knowledge. Additionally, to emphasize the variety of potential applications, we also consider an n-player coordination problem and identify when and how often do players coordinate on risk-dominant rather than payoff-dominant equilibria.[3]

Our econometric framework consists of three elements. First, we assume that the econometrician observes the joint distribution of endogenous outcomes and exogenous covariates, and knows the distribution of unobserved errors up to a finite dimensional parameter. The remaining uncertainty includes both the value of these parameters and the distribution of outcomes conditional on the observed and unobserved characteristics of the environment. We call this distribution the *distribution of play*. The second element of our framework is a *structural index* which determines the players' von Neumann Morgenstern (vNM) utilities over outcomes, and depends on the char-

[2]The experimental literature has provided some important insights as to which solution concepts may best describe behavior in laboratories (see, for instance, Camerer (2003)). However, it is not always clear whether these insights can be directly extrapolated to *all* real life situations.

[3]Equilibrium selection in coordination problems has important welfare implications in different applications including currency attacks, bank runs and regime changes (Morris and Shin, 2003), and tacit collusion in oligopolies (Green et al., 2013).

acteristics of the environment through a function known up to a finite-dimensional vector of structural parameters. Finally, in order to relate these parameters to the observed data, we introduce the notion of *solution concept*. A solution concept specifies a set of admissible distributions over outcomes depending on the value of the structural index. For example, a solution concept could specify that entry decisions must arise from NE of the game.

Our goal is to characterize which solution concepts are satisfied by actual behavior, meaning that they contain the true distribution of play. Our starting point is Theorem 2.1 in Beresteanu et al. (2011), which implies that a solution concept can rationalize the data if and only if it is satisfied *on average*, i.e., if and only if its conditional Aumann expectation[4] contains the observed distribution of choices conditional on covariates. We improve upon this characterization by establishing point identification of the structural parameters (Proposition 1.3) and the distribution of play (Proposition 1.2). By doing so, we can recover the value of the structural index and the distribution of choices conditional on both observed and unobserved heterogeneity. A solution concept is consistent with this information if and only if it is satisfied *almost surely*. This makes it possible to dispense with Aumann expectations, which are highly intractable, and to test the validity of a solution concept using, for instance, the methods explored in the companion paper **?**.

We use two assumptions to identify the distribution of play. The first one is an exclusion restriction requiring choices to be independent of *some* of the covariates, conditional on the structural index. The second one is a richness condition requiring the family of distributions of the structural index conditional on the values of the excluded covariates to be boundedly complete. The first assumption guarantees that the true distribution of play satisfies an integral equation of the first kind. The second assumption guarantees that this equation has a unique solution.[5]

The richness assumption is satisfied, for instance, if the payoffs have normally or extreme-valued distributed additive errors, as is commonly assumed in applied work. As for the exclusion restriction, one may be concerned that covariates which affect the payoffs may also affect the way people choose equilibria. However, we allow the equilibrium selection to depend both on the structural index and unobserved heterogeneity, and we only require independence conditional on the structural index.

[4]We refer the reader to Molchanov (2006) for an introduction to the theory of random sets and Aumann expectations.

[5]Complete families have been used elsewhere in the literature, for example, by Newey and Powell (2003) and Hall and Horowitz (2005) in the context of non-parametric instrumental regressions, and by Hoderlein et al. (2012) for the problem of non-parametric estimation of structural models with random coefficients. See Florens (2003) for a general discussion of invertibility of integral operators.

Hence, we still allow the excluded covariates to affect the way people choose equilibria, as long as they do so through preferences. Additionally, if there is good reason to believe that these covariates affect equilibrium selection, even after conditioning on payoffs, this hypothesis could be incorporated as part of the solution concept and tested accordingly.

We are not the first to exploit the power of completeness assumptions coupled with exclusion restrictions. In particular, Berry and Haile (2014)—as well as other related papers—apply a similar strategy to models of oligopolistic competition that allows, among other things, to discriminate between different models of competition. An important difference between their setting and ours is that they consider continuous games while we consider discrete games. In particular, the fact that the set of outcomes is uncountable allows them to relax to some extent the completeness assumption. The indispensability of some kind of completeness is an open problem in discrete settings like ours.

Our results cannot be directly applied to incomplete-information games. This is because our exclusion restriction is only consistent with solution concepts that admit selections that are independent of the excluded covariates conditional on payoffs. And solution concepts for incomplete-information games often fail to satisfy this requirement, because variables which affect the distribution of payoffs might also affect the set of equilibria. In our entry example, we still manage to identify whether payoffs are private information or common knowledge (Proposition 1.7). However, we do so relying directly on the results from Beresteanu et al. (2011), and without point identifying the distribution of play.

For the structural parameters, we establish point identification via a high-level assumption that requires different parameter values to make different predictions conditional on some set of covariate values, independently of the true solution concept. As a particular instance of this approach, we extend standard identification-at-infinity strategies (Bajari et al., 2010b, Tamer, 2003), in a way that requires only mild behavioral assumptions, namely, an extremely week notion of level-2 rationality.[6] Our general approach also encompasses the identification strategy from Kline (2015a), which does not rely on covariates with large supports.

The paper is organized as follows. Section 1.1 introduces our benchmark example, and the main potential applications that we analyze. Section 1.2 formally introduces the general econometric model. Section 1.3 deals with the identification of the distri-

[6] Aradillas-López and Tamer (2008) have already shown that, in some but not all cases, the identification power of assuming Nash play is the same as the power of assuming rationality and mutual knowledge of rationality. Our notion of rationality is slightly weaker than theirs.

bution of play, and Section 1.4 with the structural parameters. Section 1.5 presents the main results, and Section 1.6 illustrates their power and limitations with specific applications.

1.1. Motivating example: an entry game

In order to illustrate our results and suggest potential applications, we consider an entry model adapted from Bresnahan and Reiss (1990). Several papers have analyzed the identification and estimation of the payoff parameters in this kind of models, and Ciliberto and Tamer (2009) actually estimate a structural entry game using data from the airlines industry. In contrast, our focus is on identifying behavior patterns rather than payoff parameters.

Two firms $i \in I = \{1, 2\}$ must choose whether to enter a market ($y_i = 1$) or not ($y_i = 0$). Firm i's profit is given by

$$\mathbf{u}_{0i}(y) = (\beta_{0i}\mathbf{x}_i - \beta_{03}y_{-i} - \mathbf{e}_i)y_i,$$

where y_{-i} denotes the choice of i's competitor,[7] $\beta_{0i} > 0$, $i = 1, 2, 3$, are unknown fixed parameters, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ is a vector of covariates with support $X = \mathbb{R}^2$, and $\mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2)$ is a vector of unobserved error terms, independent of $\mathbf{x}$ and distributed $N(0, V_0)$. The payoff matrix for given realizations of $\mathbf{x}$ and $\mathbf{e}$ is shown in Figure 1.1. The term $\beta_{0i}x_i - e_i$ represents i's benefit or cost from entering the market, while β_{03} represents the cost of competition when both firms enter. We assume that the researcher observes the joint distribution of the covariates $\mathbf{x}$ and the entry decisions $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$.

	0	1
0	$0, 0$	$0\ ,\ \beta_{02}x_2 - e_2$
1	$\beta_{01}x_1 - e_1\ ,\ 0$	$\beta_{01}x_1 - \beta_{03} - e_1\ ,\ \beta_{02}x_2 - \beta_{03} - e_2$

Figure 1.1 – Payoff matrix for the entry game.

[7] We use a similar convention throughout the paper. Given a family $x = (x_k)_{k\in K}$ (e.g., a strategy profile, or a vectors of covariates) and a particular index value $k \in K$, we use the notation $x = (x_k, x_{-k})$ where $x_{-k} = (x_j)_{j\in K\setminus\{k\}}$.

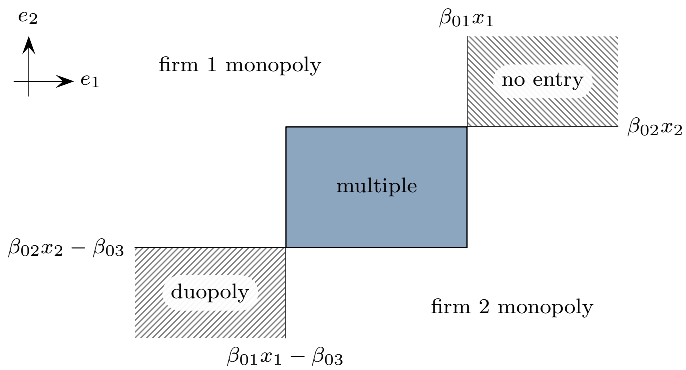


Figure 1.2 – NE correspondence for the simultaneous entry game with complete information as a function of the realization of $\mathbf{e}$.

1.1.1. Solution concepts and equilibrium selection

We consider solution concepts characterized by nonempty-valued corespondences mapping the exogenous characteristics of the environment into sets of admissible distributions for the endogenous outcomes. For example, the assumption that firms always play a NE of the simultaneous-move game with complete information can be characterized by the correspondence q_{NE} illustrated in Figure 1.2. When $e_i < \beta_{01}x_i - \beta_{03}$, entering the market is strictly dominant for player i. When $e_i > \beta_{01}x_i$, staying out of the market is strictly dominant for player i. This results in four regions of the payoff space with a unique rationalizable outcome. In each of these q_{NE} maps contains a unique admissible distribution, which assigns full probability to the each corresponding outcome. In the remaining region, q_{NE} consists of a strictly mixed NE and two pure degenerate distributions corresponding to pure strategy NE.

While dealing with this example, we maintain the assumption that firms maximize their expected profits given their information and beliefs and this fact is mutual knowledge (level-2 rationality). As we show in Section 1.4.2, this suffices to point identify the structural parameters, and allows us to focus on the task at hand.[8] We do *not* impose any further behavioral assumptions. In particular, we do *not* assume a specific game tree, that entry choices are made simultaneously or sequentially, that

[8]The identification of payoff parameters for different versions of this game has been well studied in the literature. Our approach is closest to the one in Aradillas-López and Tamer (2008) and Kline and Tamer (2012). Their analysis is also based on second-level rationality and large support of observed covariates, but they additionally assume simultaneous choices and mutual knowledge of payoffs. Tamer (2003) and Kline (2015a) characterize the sharp identified set assuming pure strategy Nash play, and without any assumptions on equilibrium selection. Henry and Mourifie (2012) find the corresponding bounds assuming Nash play in pure or mixed strategies. Bajari et al. (2010b) obtain point identification assuming Nash play, large support, and a parametric specification of equilibrium selection.

the profit functions are common knowledge, that the firms' beliefs are in equilibrium, that firms always play the same equilibrium, nor that some equilibria are more likely to arise than others. Our goal is precisely to determine from the data which of these assumptions are satisfied by the actual behavior of the firms.

An important difficulty is that the commonly used solution concepts often result in incomplete models that make set-valued predictions. Completing each of these models requires a possibly random *selection criterion* choosing elements from the random set of admissible distributions over outcomes that satisfy the solution concept. For example, it may be the case that firms randomize uniformly between the NE preferred by each of the firms. Since it is typically hard to justify a specific criterion, we make as few assumptions about it as possible. We only assume that the average strategy profile selected is conditionally independent of *some* covariates given the realized payoffs. As it turns out, as long as these covariates satisfy a richness condition, this assumption is sufficient to identify whether firm behavior complies with a solution concept or not.

1.1.2. Solution concepts and policy implications

We conclude this section with an example illustrating how assuming an incorrect solution concept may result in completely misleading policy implications. Suppose that a policymaker wants to minimize the occurrence of monopolies, and she can use a policy that would slightly decrease the competition effect (β_{03}) by a fixed amount $0 < \delta < \beta_{03}$. Further suppose that she actually knows the payoff parameters and evaluates the policy assuming that firms always play NE in pure strategies (PNE). This assumption is used, for instance, by Ciliberto and Tamer (2009). However, suppose that firms do not play NE. Their behavior complies with level-2 rationality but, they never enter when their profit functions lie in the multiplicity region. This could happen, for instance, if the firms were ambiguity averse when facing strategic uncertainty, and they used maxmin strategies when the game is not dominance solvable.

Pure equilibria in the multiplicity region always predict monopolies. Hence, under the PNE hypothesis, monopolies arise in all payoff regions except for those in which either not entering is dominant for both firms, or entering is dominant for both firms. The policy being evaluated increases the probability of the latter region (see Figure 1.3). Therefore, under the policymaker's assumptions, the policy unambiguously reduces the occurrence of monopolies independently of the parameter values.

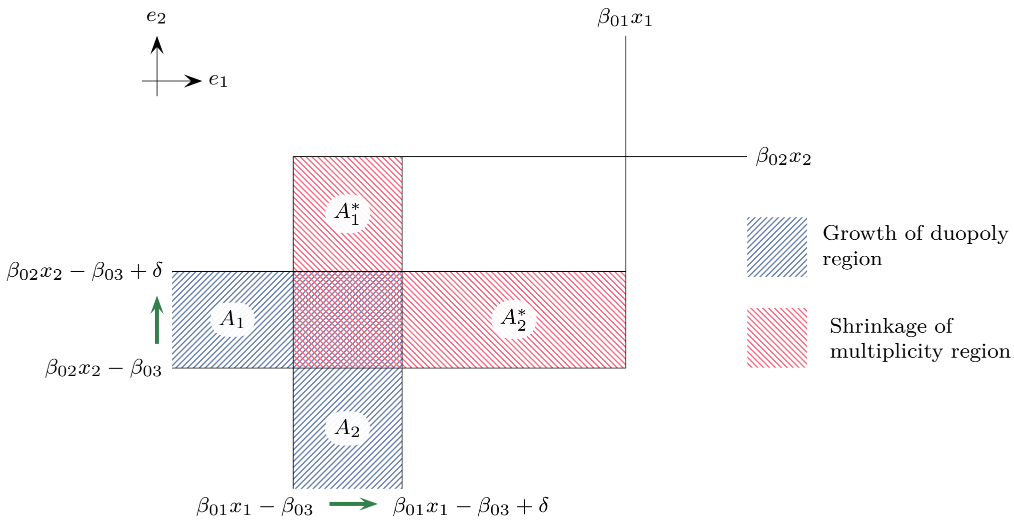


Figure 1.3 – Effect of the proposed policy.

However, given the true behavior of the firms, the effect of the policy is always smaller than under the PNE hypothesis, and it can even have the opposite direction for some parameter values. This happens because the policy also decreases the probability of the multiplicity region and, under the true behavior, monopolies never arise in this region. The net effect of the policy on the probability of monopolies is given by the difference between the probability of regions $(A_1 \cup A_2)$ and $(A_1^* \cup A_2^*)$ in Figure 1.3. For some parameter values and realized covariate values the adverse effect can actually dominate, and the policy may actually increase the probability of a monopoly. For example, one can easily verify that this is the case whenever the error terms are i.i.d. standard normal and $1 \leq \beta_{01}x_1 = \beta_{02}x_2 \leq (\beta_{03} - \delta)/2$.

1.2. Econometric framework

1.2.1. Data generating process

Each instance of the environment is characterized by an endogenous outcome y from a finite set Y, a vector of exogenous characteristics observed by the econometrician $x \in X \subseteq \mathbb{R}^{d_x}$, and a vector of unobserved characteristics $e \in E \subseteq \mathbb{R}^{d_e}$.

Assumption 1.1 (Data generating process) Random objects $\mathbf{y} : \Omega \to Y$, $\mathbf{x} : \Omega \to X$, and $\mathbf{e} : \Omega \to E$, are defined on a probability space $(\Omega, \mathscr{F}, \Pr)$. The researcher observes the joint distribution of $(\mathbf{x}, \mathbf{y})$ and knows the distribution of $\mathbf{e}$ conditional on $\mathbf{x}$ up to a finite-dimensional vector of unknown parameters γ_0 belonging to a known compact set $\Gamma \subseteq \mathbb{R}^{d_\gamma}$.

Remark 1.1 For the rest of the paper, we assume that Γ is a singleton, i.e., γ_0 is known. This is a normalization that comes without loss of generality, because e only enters our structural model through a function which is also unknown up to a finite-dimensional parameter. Hence, the uncertainty about γ_0 can be embedded into the parameters describing this function. Our analysis extends straightforwardly to the case in which Γ_0 is not a singleton, as long as it is finite dimensional. For example, in the entry game, the variance of $\mathbf{e}$ can be normalized to be the identity matrix, by adding an additional matrix coefficient to the function mapping $(\mathbf{e}, \mathbf{x})$ into $\mathbf{u}_0$.

With slight abuse of notation, we identify distributions over Y with vectors on the $\|Y\|$-dimensional simplex $\Delta(Y)$. We denote the observed distribution of $\mathbf{y}$ conditional on $\mathbf{x}$ by $\mu_0(\mathbf{x})$, and the unknown distribution of $\mathbf{y}$ conditional on $\mathbf{x}$ and $\mathbf{e}$ by $h_0(\mathbf{e}, \mathbf{x})$. Let H be the set of measurable functions from $E \times X$ to $\Delta(Y)$. We call each such function a possible *distribution of play*, and h_0 the true distribution of play. We use the notation $\mathbf{h} = h(\mathbf{e}, \mathbf{x})$, and $\mathbf{h}_0 = h_0(\mathbf{e}, \mathbf{x})$. A possible distribution of play $h \in H$ is consistent with the observed data if and only if:

$$\mu_0(\mathbf{x}) = \mathbb{E}[\,\mathbf{h}|\mathbf{x}\,] \quad \text{a.s.}. \tag{1.1}$$

By construction, we know that h_0 is consistent with the observed data.

Remark 1.2 [Notation] Note that h is a function mapping $E \times X$ into $\Delta(Y)$, while $\mathbf{h}$ is a function mapping Ω into $\Delta(Y)$. More specifically, $\mathbf{h}$ is the random vector in $\Delta(Y)$ obtained from the composition of the deterministic function h and the random vector $(\mathbf{e}, \mathbf{x})$. We use a similar notation for the structural index and solution concepts defined ahead, and the best response correspondences introduced in Section 1.4.1.

Example 1.1 Consider our entry model. Suppose that agents always randomize uniformly across all NE. For (e, x) in the multiplicity region, let $p^*_{-i} = (\beta_{i0}x_i - e_i)/\beta_{03}$ be the probability of firm $-i$ entering in the unique mixed NE (for brevity of notation we suppress the explicit dependence on (e, x)). Then, the true distribution of play is characterized (up to a null-set) by:

$$h_0(e,x) = \begin{cases} \delta_{(1,1)} & \text{if} \quad e_i < \beta_{0i}x_i - \beta_{03} \text{ for } i = 1,2 \\ \delta_{(0,0)} & \text{if} \quad e_i > \beta_{0i}x_i \text{ for } i = 1,2 \\ \delta_{(1,0)} & \text{if} \quad e \in E_1(x) \\ \delta_{(0,1)} & \text{if} \quad e \in E_2(x) \\ \psi^*(e,x) & \text{otherwise} \end{cases},$$

where δ_y denotes the distribution that assigns full probability to outcome y,

$$E_i(x) = \left\{ e \left| \begin{array}{c} (e_i < \beta_{0i}x_i - \beta_{03} \quad \text{and} \quad e_{-i} > \beta_{0-i}x_{-i} - \beta_{03}) \\ \text{or} \quad (e_i < \beta_{0i}x_i \quad \text{and} \quad e_{-i} > \beta_{0-i}x_{-i}) \end{array} \right. \right\}$$

corresponds to the region where i's monopoly is the only rationalizable outcome for $i = 1,2$ (see Figure 1.2), and

$$\psi^*(e,x) = \frac{1}{3}\Big((1-p_1^*)(1-p_2^*),\, 1+p_2^*(1-p_1^*),\, 1+p_1^*(1-p_2^*),\, p_1^*p_2^*\Big)$$

specifies the probability of each outcome $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$, respectively, in the multiplicity region.

1.2.2. Structural parameters

The researcher is interested in a random structural index which takes values in a known set $U \subseteq \mathbb{R}^{d_u}$, and is given by a parametric function of the primitive exogenous characteristics of the environment. To fix ideas, we let $U \subseteq \mathbb{R}^{\|I\| \times \|Y\|}$, and we interpret the value of index as representing the agent's preferences over outcomes. In different applications, it could receive alternative interpretations. For example, it could contain information about beliefs or best response correspondences.

Assumption 1.2 (Structural parameters) The structural index $\mathbf{u}_0 : \Omega \to U$ is given by $\mathbf{u}_0 = g(\beta_0, \mathbf{e}, \mathbf{x})$, where $g : B \times E \times X \to U$ is a known measurable function and β_0 is a finite-dimensional vector of parameters belonging to a known set $B \subseteq \mathbb{R}^{d_\beta}$. We use the notation $\mathbf{u}_\beta = g(\beta, \mathbf{e}, \mathbf{x})$ for $\beta \in B$, so that $\mathbf{u}_0 = \mathbf{u}_{\beta_0}$.

We have not yet introduced any link connecting β_0 to the observed data described in Assumption 1.1. Hence, we need additional assumptions, both in order to learn something about β_0 from the data, and to make predictions based on each candidate value of β_0.

1.2.3. Solution concepts

We focus on a special kind of structural assumptions which make set predictions about the distribution of endogenous outcomes as a function of the characteristics of the environment.

Definition 1.1 (Solution concept) A *solution concept* is a nonempty-valued and closed-valued correspondence $q : B \times E \times X \rightrightarrows \Delta(Y)$, such that

$$\left\{\omega \in \Omega \,\middle|\, \quad \Psi \cap q(\beta, \mathbf{e}(\omega), \mathbf{x}(\omega)) \neq \emptyset\right\} \in \mathscr{F}, \tag{1.2}$$

for every $\beta \in B$ and every *closed* set $\Psi \subseteq \Delta(Y)$.

We use the notation $\mathbf{q}(\beta) = q(\beta, \mathbf{e}, \mathbf{x})$. Equation (1.2) is a measurability condition requiring that $\mathbf{q}(\beta)$ should be a random set for every $\beta \in B$ (Molchanov, 2006). Although we define solution concepts in terms of primitives $\mathbf{x}$ and $\mathbf{e}$, and not in terms of $\mathbf{u}_0$, the fact that they depend on the structural parameters provides the missing link between the data and β_0.

Definition 1.2 A pair $(\beta, h) \in B \times H$ *jointly satisfies* q if:

$$\mathbf{h} \in \mathbf{q}(\beta) \quad \text{a.s.}, \tag{1.3}$$

and q is *satisfied* if (β_0, h_0) jointly satisfies it.

Example 1.2 We have already provided q_{NE} as an example of a solution concept for the entry game. However, there is a remark to be made. The statement $\mathbf{h}_0 \in \mathbf{q}_{\mathrm{NE}}$ implies that the average distribution of choices conditional on payoffs corresponds to a particular NE. Hence, it assumes implicitly that agents always play the same equilibrium given $(\mathbf{e}, \mathbf{x})$. For instance, it is not satisfied by the distribution of play from Example 1.1. To dispense with this implicit assumption, we have to consider the possibility that the actual distribution of play is a *mixture* of equilibria and could thus correspond to any point in the convex hull of $\mathbf{q}_{\mathrm{NE}}$. We accommodate this possibility by considering instead the solution concept $\bar{q}_{\mathrm{NE}}(\beta, e, x) = \mathrm{co}(q_{\mathrm{NE}}(\beta, e, x))$, where $\mathrm{co}(\cdot)$

denotes the convex-hull operator.

1.2.4. Identification

The main objective of the paper is to determine which solution concepts accurately characterize the behavior of the agents, that is, which solution concepts are satisfied. To address this issue, we have to determine which distributions of play and structural parameters are consistent with the data. And to narrow these sets, we restrict the parameter space through a series of structural assumptions, each of them stating that (β_0, h_0) belongs to some known set $\Upsilon \subseteq B \times H$.

Definition 1.3 (Sharp identified set) The *sharp identified set* for (β_0, h_0) under structural assumptions $\{\Upsilon_l\}_{l=1}^{L}$ consists of the set of pairs $(\beta, h) \in B \times H$ such that $(\beta, h) \in \Upsilon_k$ for $l = 1, \ldots, L$, and h is consistent with the observed data, i.e., it satisfies (1.1).

Definition 1.4 A solution concept q is *consistent with the data* under structural assumptions $\{\Upsilon_l\}_{l=1}^{L}$ if it is jointly satisfied by some (β, h) belonging to the sharp identified set under $\{\Upsilon_l\}_{l=1}^{L}$.

We have thus defined two different properties for any given solution concept. We say that it is satisfied if the actual distribution of play belongs to it almost surely, and we say that it is consistent if the observed data could be generated by a choice pattern that satisfies it. Consistency can be directly tested from the data at least in some cases (see, for instance, **?**). However, the researcher may be interested in whether a solution concept accurately characterizes behavior, as this may be crucial for the validity of counterfactual analyses. Our main result (Theorem 1.5) establishes conditions under which these two properties are equivalent, thus enabling to test which solution concepts are satisfied.

As the first step in characterizing the set of consistent solution concepts, applying Theorem 2.1 in Beresteanu et al. (2011) to our framework yields the following result.

Theorem 1.1 (Beresteanu et al. (2011)) *Under assumptions 1.1 and 1.2, a solution concept q is consistent with the data if and only if there exists some $\beta \in B$ such that*

$$\mu_0(\mathbf{x}) \in \mathbb{E}[\,\mathbf{q}(\beta)|\mathbf{x}\,] \quad \text{a.s.,}$$

where $\mathbb{E}[\,\mathbf{q}(\beta)|\mathbf{x}\,]$ *denotes the conditional Aumann expectation of* $\mathbf{q}$*.*

The conditional Aumann expectation is taken over the unobserved heterogeneity embedded in $\mathbf{e}$. Hence, according to this result, a solution concept is consistent as long as its restrictions are satisfied *on average*, conditioning only on observed covariates. In contrast, under our assumptions, they should be satisfied *ex post*, conditioning on both observed covariates and unobserved error terms. This matters not only for interpretation purposes, but also because it allows to dispense with Aumann expectations, which are often computationally intractable. The key step in our analysis is to establish point identification of β_0 and h_0. We do in the following sections.

1.3. Identifying the distribution of play

The researcher observes the distribution of outcomes conditional on $\mathbf{x}$. That is why Aumann expectations appear in Theorem 1.1. In this section, we make two assumptions that allow us to solve Equation (1.1), in order to recover the distribution of outcomes conditional on both $\mathbf{x}$ and $\mathbf{e}$, i.e., the true distribution of play. The first assumption is an exclusion restriction requiring that some of the observed covariates only affect choices through the structural index. The second assumption is a richness condition requiring that these covariates have sufficient heterogeneity and generate sufficient variation in the distribution of the structural index.

Suppose that the observable covariates can be written as $\mathbf{x} = (\mathbf{w}, \mathbf{z})$, where $\mathbf{z}$ represents those excluded covariates that only affect choices through the structural index. Let W and Z denote the supports of $\mathbf{w}$ and $\mathbf{z}$, respectively.

Assumption 1.3 (Exclusion restriction) The distribution of play h_0 is measurable with respect to the σ-algebra generated by $(\mathbf{u}_0, \mathbf{w})$, i.e., there exists a measurable function $\tilde{h}_0 : U \times W \to \Delta(Y)$ such that $h_0(\mathbf{e}, \mathbf{x}) = \tilde{h}_0\big(\mathbf{u}_0, \mathbf{w}\big)$ a.s..

Remark 1.3 Assumption 1.3 restricts the set of solution concepts that can be satisfied. For a solution concept q to be consistent with this assumption, it has to admit a selection that is measurable with respect to $(\mathbf{u}_0, \mathbf{w})$. In our entry example, this rules out the possibility that firms play BNE of the incomplete information game, because the set of BNE changes depending on the public information contained in $\mathbf{x}$. See Section 1.6.5.

In terms of the underlying game, and having a fixed solution concept in mind, Assumption 1.3 imposes some restrictions on the selection criterion. Namely, it requires the *average* of the selected equilibria to be independent of $\mathbf{e}$ and $\mathbf{z}$ conditional of $\mathbf{w}$ and $\mathbf{u}_0$. This is the only assumption that we impose on selection criteria. Also, by only assuming conditional independence, we allow for the possibility that $\mathbf{z}$ could affect the selection, but only through $\mathbf{u}_0$ or $\mathbf{w}$. This assumption is commonly used in the literature, although it is often not explicitly assumed.[9] For instance, Bajari et al. (2010b) allows equilibrium selection to be random, but the probability of choosing each equilibrium is fully parametric and depends only on payoffs. In turn, other papers make the much stronger assumption that players always choose the same equilibrium.

Assumption 1.3 guarantees that h_0 does not depend on z, given u and w. Hence, the consistency condition (1.1) becomes an integral equation of the first kind. In order to ensure that it has a unique solution, we require $\mathbf{z}$ to generate enough variation in the distribution of the structural index. We do so by restricting the parameter space for β, so as to guarantee that the collection of distributions of $\mathbf{u}_\beta$ conditional on different realizations of $\mathbf{z}$ constitutes a boundedly complete family.[10]

Assumption 1.4 (Bounded completeness) For every measurable and bounded function $r : U \times W \to [0, 1]$, if $\mathbb{E}[\, r(\mathbf{u}_{\beta_0}, \mathbf{w})|\mathbf{w}, \mathbf{z}\,] = 0$ a.s., then $r(\mathbf{u}_{\beta_0}, \mathbf{w}) = 0$ a.s..

Assumption 1.4 is implied by the following parametric assumption, which is satisfied by all models with additive errors with normal or extreme-valued distributions, and by more general specifications commonly used both in applied work (e.g., Ciliberto and Tamer (2009)), and in numerical simulations from theoretical work (e.g., Kline (2015a)).

Assumption 4' There exists a random variable $\mathbf{v}$ with support $V \subseteq \mathbb{R}^{d_v}$, and known functions $f : B \times V \times W \to U$ and $m : Z \to \mathbb{R}^{d_v}$ such that (i) $\mathbf{u}_\beta = f(\beta, \mathbf{v}, \mathbf{w})$ a.s., (ii) the distribution of $\mathbf{v}$ conditional on $\mathbf{x}$ belongs to the exponential family with $m(\mathbf{z})$

[9] Some exceptions include Tamer (2003), Beresteanu et al. (2011) and Kline (2015a), but they rely on ruling out mixed strategies to obtain point identification. In fact, Henry and Mourifie (2012) shows that non-parametric point identification of payoffs assuming Nash equilibrium is not possible in 2×2 games without imposing restrictions on the selection criteria.

[10] We refer the reader to Andrews (2011) for more details on boundedly complete families.

as a parameter, and (iii) the support of $m(\mathbf{z})$ conditional on $\mathbf{w}$ contains an open set a.s..

Condition (i) in Assumption 4' means that $\mathbf{v}$ is a sufficient statistic for $\mathbf{u}_\beta$ given $\mathbf{w}$. Conditions (ii) and (iii) are standard assumptions which imply that the collection of distributions of $\mathbf{v}$ conditional on different realizations of $\mathbf{z}$ given β and $\mathbf{w}$ constitutes a complete family of distributions, see, for instance, Theorem 2.12 in Brown (1986). Since $\mathbf{u}_\beta$ is measurable with respect to $\mathbf{v}$, the collection of conditional distributions of $\mathbf{u}_\beta$ is also complete. Therefore, Assumption 4' implies Assumption 1.4.

There are two crucial restrictions imposed by Assumption 4'. First, the distribution of errors should belong to the exponential family (or another boundedly complete family).[11] Second, there must be continuous error-specific covariates that shift the effect of each error term. It is very flexible in terms of the functional forms that satisfy it. It is satisfied by our entry game (see Section 1.3.1), and by pricing games both with linear and logistic demand systems as the ones studied in Berry et al. (1995). It is also satisfied when v is a linear function of e and x and the covariate space has at least the same dimension as the error space, and therefore it incorporates a large class of multiple-index models. Consider for instance the following example based on the classical teamwork model from Hölmstrom (1982).

Example 1.3 Partners $i \in I = \{1, \ldots, \iota\}$ participate in a joint-venture. Each partner provides a level of non-contractible effort $y_i \in \{0, 1\}$. The private cost $\mathbf{c}_i$ and productivity $\mathbf{a}_i$ for agent i are given by $\mathbf{c}_i = c(\beta, \mathbf{w}, \beta_{ic}\mathbf{z}_{ic} + \mathbf{e}_{ic})$ and $\mathbf{a}_i = a(\beta, \mathbf{w}, \beta_{ia}\mathbf{z}_{ia} + \mathbf{e}_{ia})$, where $(\mathbf{e}_{ic}, \mathbf{e}_{ia})$ are agent specific shocks, $(\mathbf{z}_{ic}, \mathbf{z}_{ia})$ are excluded covariates, and c and a are known functions. *Per capita* output is given by a known function π of $\mathbf{a} = (\mathbf{a}_i)_{i \in I}$ and $y = (y_i)_{i \in I}$. Output is shared equally so that the realized utility function for player i is $\mathbf{u}_{\beta i}(y) = \pi(\mathbf{a}, y) - y_i \mathbf{c}_i$

Assume that the agent specific shocks $(\mathbf{e}_{ic}, \mathbf{e}_{ia})_{i \in I}$ are independent and identically distributed according to the Gumbell distribution with parameter $(0, 1)$, that is, the c.d.f. of each $\mathbf{e}_{ij}$ is $F_{\mathbf{e}_{ij}}(e_{ij}) = \exp(-\exp(-e_{ij}))$. Let $\mathbf{v} = (\beta_{ic}\mathbf{z}_{ic} + \mathbf{e}_{ic}, \beta_{ia}\mathbf{z}_{ia} + \mathbf{e}_{ia})_{i \in I}$ and $\mathbf{z} = (\mathbf{z}_{ic}, \mathbf{z}_{ia})_{i \in I}$. Clearly, $\mathbf{u}_\beta$ can be expressed as a known function of $(\mathbf{w}, \mathbf{v})$. Furthermore, it can be shown that the family of distributions of $\mathbf{v}$ conditional on $\mathbf{x} = (\mathbf{w}, \mathbf{z})$ belong to the exponential family with $m(\mathbf{z}) = (\beta_{ic}\mathbf{z}_{ic}, \beta_{ia}\mathbf{z}_{ia})_{i \in I}$ as a parameter.[12] Hence, Assumption 4' is satisfied as long as the support of $\mathbf{z}$ conditional

[11] Unfortunately, testing whether the error term distribution belongs to a complete family is not possible in general (Canay et al., 2013, Freyberger, 2014).

[12] Indeed, let $\tilde{v}_i = \beta_{ic}z_{ic} + e_{ic}$, $\tilde{v}_{2i} = \beta_{ia}z_{ia} + e_{ic}$, $\tilde{m}_i = \beta_{ic}z_{ic}$ and $\tilde{m}_{2i} = \beta_{ia}z_{ia}$ for $i = 1, \ldots, \iota$, so

on $\mathbf{w}$ contains an open ball and $\beta_{ic}, \beta_{ia} \neq 0$ for all i.

Assumptions 1.3 and 1.4 are sufficient to guarantee that the true distribution of play is point identified up to the vector of unknown structural parameters β_0.

Proposition 1.2 *Under assumptions 1.1, 1.2, 1.3 and 1.4, if (β, h) and (β, h') belong to the sharp identified set then $\mathbf{h} = \mathbf{h}'$ a.s..*

Proof. Let (β, h) and (β, h') be consistent with the data and satisfy the assumptions of the proposition. The consistency condition (1.1) implies that

$$\mathbb{E}[\,\mathbf{h} - \mathbf{h}'|\mathbf{x}\,] = \mathbb{E}[\,\mathbf{h}|\mathbf{x}\,] - \mathbb{E}[\,\mathbf{h}'|\mathbf{x}\,] = \mu_0(\mathbf{x}) - \mu_0(\mathbf{x}) = 0 \quad \text{a.s..}$$

By Assumption 1.3, there exist $\tilde{h}, \tilde{h}' : U \times W \to \Delta(Y)$ such that $\mathbf{h} = \tilde{h}\Big(\mathbf{u}_\beta, \mathbf{w}\Big)$ and $\mathbf{h}' = \tilde{h}'\Big(\mathbf{u}_{\beta'}, \mathbf{w}\Big)$ a.s.. Therefore, we have that

$$\mathbb{E}\Big[\,\tilde{h}(\mathbf{u}_\beta, \mathbf{w}) - \tilde{h}'(\mathbf{u}_{\beta'}, \mathbf{w})|\mathbf{x}\Big] = \mathbb{E}[\,\mathbf{h} - \mathbf{h}'|\mathbf{x}\,] = 0 \quad \text{a.s..}$$

Because β satisfies Assumption 1.4, it follows that $\tilde{h}(\mathbf{u}_\beta, \mathbf{w}) - \tilde{h}'(\mathbf{u}_{\beta'}, \mathbf{w})$ a.s., and, consequently, $\mathbf{h} = \mathbf{h}'$ a.s.. ■

1.3.1. Distribution of play in the entry game

Now let us consider the problem of identifying the distribution of play in our entry model. As we discussed previously, Assumption 1.3 cannot be satisfied when the firms play BNE of the incomplete information game. Hence, we assume that profit functions are common knowledge and the true solution concept is measurable with respect to $\mathbf{u}_0$. Under these conditions, Assumption 1.3 is satisfied as long as the selection criterion is independent of $\mathbf{x}$ conditional on $\mathbf{u}$, that is, whenever it depends

that $v = (\tilde{v}_i, \tilde{v}_{2i})_{i \in I}$ and $m(z) = (\tilde{m}_i, \tilde{m}_{2i})_{i \in I}$. Then, the p.d.f. of $\mathbf{v}$ conditional on $\mathbf{x}$ is:

$$\begin{aligned} f_{\mathbf{v}|\mathbf{x}}(v|x) &= \prod_{i=1}^{\iota} f_{\mathbf{e}_{ic}}(v_{ic} - \beta_{ic} z_{ic}) f_{\mathbf{e}_{ia}}(v_{ia} - \beta_{ia} z_{ia}) \\ &= \Psi(v) \exp\left(\sum_{i=1}^{2\iota} \eta_i(m(z)) T_i(v) - \phi(m(z))\right), \end{aligned}$$

where $\Psi(v) = \exp\left(-\sum_{i=1}^{\iota}(\tilde{v}_i + \tilde{v}_{2i})\right)$, $\phi(m(z)) = \sum_{i=1}^{\iota}(\tilde{m}_i + \tilde{m}_{2i})$, and $T_i(v) = \exp(-\tilde{v}_i)$, $T_{2i}(v) = \exp(-\tilde{v}_{2i})$, $\eta_i(m(z)) = \tilde{m}_i$ and $\eta_{2i}(m(z)) = \tilde{m}_{2i}$, for $i = 1, \ldots, \iota$.

only on the firms' preferences and unobserved heterogeneity.

There could be some observable covariates that affect the selection criteria beyond their direct effect on payoffs. These may include, for example, the relative size of the firms, the relationship of the firms with the market (domestic vs. foreign), or the experience of the firms' managers. Thankfully, Assumption 1.3 does not require that the selection should be independent of *all* covariates. It could still be satisfied as long as $\mathbf{h}_0$ is independent of *some* covariates $\mathbf{z}$ which generate enough variation in $\mathbf{u}_0$.

As for Assumption 1.4, note that $\mathbf{u}_\beta$ can be written as $\mathbf{u}_i(y) = (\mathbf{v}_i - \beta_{03}y_{-i})y_i$, where $\mathbf{v}_i = \beta_{0i}\mathbf{x}_i - \mathbf{e}_i$. Note that

$$\mathbf{v}|\mathbf{x} = x \sim N\left(\begin{pmatrix}\beta_{10}x_1 \\ \beta_{20}x_2\end{pmatrix}, V\right).$$

Since normal distributions belong to the exponential family, Assumption 4' is satisfied. Hence, the true distribution of play is point identified up to β_0 as long as we assume that the true solution concept and selection criterion are independent of $\mathbf{x}$ conditional on $\mathbf{u}_0$.

1.4. Point identification of structural parameters

It may be important to identify solution concepts in settings in which the structural parameters are only partially identified. However, to keep the discussion focused, we impose a high-level assumption that is sufficient for point identification of β_0. Our assumption encompasses standard approaches used elsewhere in the literature, including, for instance, the identification-at-infinity approach from Bajari et al. (2010b) and the identification strategy with bounded covariates from Kline (2015a).[13] The rest of this section provides more tractable lower-level assumptions and establishes point identification of β_0 in our entry model.

[13] Kline (2015a)'s identification strategy for entry games is very attractive in that it does not require covariates with large supports. However, it is not well suited for our purposes because it relies on a strong assumption on the solution concept. Namely, he assumes that firms always play NE in *pure* strategies. Still, his assumptions are encompassed by our Assumption 1.2. First, assuming pure NE implies that the only possible outcomes in the multiplicity region are $(1,0)$ and $(0,1)$. Therefore, the conditional probabilities of $(0,0)$ and $(1,1)$ are known functions of β and x, and our condition (1.4) is satisfied. Furthermore, the assumptions that he uses to guarantee that these probabilities identify the payoff parameters guarantee that our condition (1.5) is also satisfied.

Assumption 5 (Identifying single predictions) There exists a known family of tuples $(h_k, Y_k, (X_k^n)_{n=1}^{\infty})_{k\in K}$, with each tuple consisting of a measurable function $h_k : B \times E \times X \to \Delta(Y)$, a set of outcomes $Y_k \subseteq Y$, and a sequence of sets of covariate values $X_k^n \subseteq X$ with $\Pr(X_k^n) > 0$, such that for every k:

$$\lim_{n\to\infty} \Pr\left(\Big[h_0(\mathbf{e},\mathbf{x})\Big](Y_k) = \Big[h_k(\beta_0,\mathbf{e},\mathbf{x})\Big](Y_k)\,\Big|\,\mathbf{x}\in X_k^n\right) = 1, \tag{1.4}$$

and for every $\beta \in B \setminus \{\beta_0\}$ there exists some k such that:

$$\lim_{n\to\infty} \mathbb{E}\left[\Big[h_k(\beta_0,\mathbf{e},\mathbf{x})\Big](Y_k) - \Big[h_k(\beta,\mathbf{e},\mathbf{x})\Big](Y_k)\,\Big|\,\mathbf{x}\in X_k^n\right] \neq 0. \tag{1.5}$$

Condition (1.4) requires that, for some (limiting) regions of the covariates' support, the conditional probability of some outcomes are known up the structural parameter. Condition (1.5) then requires that these probabilities identify β_0.

Proposition 1.3 *Under assumptions 1.1, 1.2, and 5, β_0 is point identified.*

Proof. If (β, h) belongs to the identified set under Assumption (5), Condition (1.4) implies that

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}\left[\Big[h_k(\beta,\mathbf{e},\mathbf{x})\Big](Y_k)\,\Big|\,\mathbf{x}\in X_k^n\right] &= \lim_{n\to\infty} \mathbb{E}\left[\Big[h(\mathbf{e},\mathbf{x})\Big](Y_k)\,\Big|\,\mathbf{x}\in X_k^n\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\Big[h_0(\mathbf{e},\mathbf{x})\Big](Y_k)\,\Big|\,\mathbf{x}\in X_k^n\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\Big[h_k(\beta_0,\mathbf{e},\mathbf{x})\Big](Y_k)\,\Big|\,\mathbf{x}\in X_k^n\right],
\end{aligned}$$

for all $k = 1, \ldots, K$, where the second equality follows from (1.1). Moreover, if Assumption (5) is satisfied and $\beta \neq \beta_0$, then, by Condition (1.5), there would exist some k such that

$$\lim_{n\to\infty} \mathbb{E}\left[\Big[h_k(\beta_0,\mathbf{e},\mathbf{x})\Big](Y_k)\,\Big|\,\mathbf{x}\in X_k^n\right] \neq \lim_{n\to\infty} \mathbb{E}\left[\Big[h_k(\beta,\mathbf{e},\mathbf{x})\Big](Y_k)\,\Big|\,\mathbf{x}\in X_k^n\right].$$

Therefore, if (h, β) belongs to the sharp identified set under assumptions 1.1, 1.2, and 5, then $\beta = \beta_0$. ■

Remark 1.4 We conjecture that Assumption 5 is almost necessary for point identification of the structural parameters. It is not necessary because point identification is sometimes possible even when all known predictions h_k are set valued. In such cases, Condition (1.5) should be reformulated to require their Aumann expectations to be disjoint. For an example in which it is not necessary, see the discussion following Proposition 1.4.

1.4.1. Identification at infinity

The standard identification-at-infinity approach consists of assuming equilibrium play—which refines rationalizable play—and the existence of covariates that make the game dominance-solvable in two steps at the limit. Our assumptions are slightly more general in order to accommodate existing solution concepts which do not imply Bayesian rationality nor rationalizability,[14] but the logic is the same. We assume that, for some values of some covariates, some players make choices to solve a given optimization problem with probability approaching one (Assumption 6), and that the corresponding choice models are identified (Assumption 7).

Formulating these assumptions requires some additional structure. Suppose that the set of observable outcomes can be written as a profile of players' actions. That is, let $Y = \times_{i \in I} Y_i$, where Y_i is the set of outcome characteristics that player i controls. In our entry model we have $Y_i = \{0, 1\}$.

Since we are neither specifying a particular game tree nor a particular information structure, elements of Y_i should be interpreted as actions rather than strategies. However, we still can define best response correspondences, *as if* Y_i was a set of strategies in a simultaneous move game. Let

$$\text{BR}_i(y_{-i}; \beta, e, x) = \arg\max_{y_i \in Y_i} \Big\{ [g_i(\beta, e, x)](y_i, y_{-i}) \Big\}$$

be the set of actions $y_i \in Y_i$ that would maximize i's preferences given e and x if i's opponents chose y_{-i}, and the value of true structural parameter was β. Also, let $\mathbf{BR}_i(y_{-i}; \beta) = \text{BR}_i(y_{-i}; \beta, \mathbf{e}, \mathbf{x})$ be the random set of best responses to y_{-i} given β.

Assumption 6 For every player i and each $y_{-i} \in Y_{-i}$, there exists a covariate $\mathbf{x}_k$ such that (i) $\mathbf{u}_{0i}$ and $\mathbf{x}_k$ are independent conditional on $\mathbf{x}_{-k} = (\mathbf{x}_j)_{j \neq k}$, (ii) the support of

[14] For example, an agent who exhibits ambiguity-aversion may find it optimal to play a strategy with a good worst-case scenario even though it may be dominated by a mixed strategy. Alternatively, Bayesian players could choose dominated strategies according to different far-sighted equilibrium concepts based on implicit or explicit forms of commitment or repetition, as in a repeated prisoners' dilemma.

$\mathbf{x}_{-k}$ is independent of $\mathbf{x}_k$, (iii) the support of $\mathbf{x}_k$ is unbounded from above, and (iv):

$$\lim_{x_k \to \infty} \mathbb{E}\left[\mathbf{h}_0\Big(\mathbf{BR}_i(y_{-i};\beta_0)\Big) \,\Big|\, \mathbf{x}_k \geq x_k \right] = 1. \tag{1.6}$$

Assumption 6 is a joint assumption on the distribution of play and the payoff distribution. Conditions (iii) and (iv) essentially guarantee that, for large values of $\mathbf{x}_k$, player i faces a single agent decision problem without uncertainty, and makes choices which maximize his utility, under every h under consideration. Conditions (i) and (ii) guarantee that $\mathbf{x}_k$ does not affect the distribution of i's payoffs. These conditions jointly identify the marginal distribution of best responses to y_{-i}.

Remark 1.5 Assuming that there is one such covariate for every i and y_{-i} may be unrealistic practice, specially for large games. We maintain the assumption for exposition purposes. In reality, we may need this to hold for only a few action profiles. For example, in our entry game, we only need this to hold for two action profiles. Also, the identified set for β_0 can be largely reduced even if there are insufficient restrictions to obtain point identification.

In order to further understand Assumption 6 and the kind of environments in which it is satisfied, one may think of it as implicitly containing two separate assumptions. The first implicit assumption is that, when the value of $\mathbf{x}_k$ is high, the payoffs that players $-i$ obtain from playing y_{-i} become arbitrarily high, and hence they are willing to play it for sure. This part of the assumption can be guaranteed, for instance, if the following two conditions are met. The first condition—on the distribution of payoffs—is that we can write:

$$\mathbf{u}_{\beta,-i}(y) = \mathbf{v}_{-i}(y;\beta) + \mathbf{x}_k \cdot \mathbb{1}(y_{-i} = y^*_{-i}),$$

where $\mathbf{v}_{-i}$ is independent of $\mathbf{x}_k$.

The second condition—on the set of distributions of play—is that there exists some $M \geq 1$ such that:

$$\min_{y_{-i} \in Y_{-i}} \mathbf{u}_{0i}(y^*_i, y_{-i}) > M \cdot \max_{y_i \neq y^*_i} \max_{y_{-i} \in Y_{-i}} \mathbf{u}_{0i}(y) \quad \Rightarrow \quad \mathbf{h}_0(\{y^*_i\} \times Y_{-i}) = 1 \quad \text{a.s..}$$

That is, if the worst payoff that player i can get from playing y^*_i is at least M times bigger than the best payoff she can get from playing anything else, then she plays y^*_i for sure.[15] Note that this restriction is considerably weaker than Bayesian rationality.

[15]This condition is closely related to the notions of counterfactual rationalizability and abso-

It is satisfied by all solution concepts which imply (myopic) rationality, and by those which imply individual rationality. In particular, it is satisfied by all of the solution concepts considered in our examples.

This first implicit assumption guarantees that, on the limit, players $-i$ choose y^*_{-i} with probability approaching one. The second implicit assumption is that, in such extreme cases in which there is no strategic uncertainty, player i chooses actions which maximize her utility. Hence, we do assume Bayesian rationality, but only at limit points in which the environment can be thought of as a single-agent decision problem without uncertainty. In any case, our assumptions are implied by standard assumptions used elsewhere in the literature, for instance, in Bajari et al. (2010b).

Assumption 6 guarantees that, asymptotically, players faces single-agent decision problems. The following step is to assume that the observed distribution of optimal choices in these problems identifies the parameters that govern the *marginal* distribution of i's payoffs.[16]

Assumption 7 For every $\beta \in B \setminus \{\beta_0\}$ there exists a player i, a strategy profile y^*, and a set $X' \subseteq X$, such that $\Pr(X') > 0$ and at least one of the following two conditions hold,

$$\Pr\left(\mathbf{BR}_i(y^*_{-i};\beta_0) = \{y^*_i\} \,\middle|\, \mathbf{x} \in X'\right) > \Pr\left(\mathbf{BR}_i(y^*_{-i};\beta) \ni y^*_i \,\middle|\, \mathbf{x} \in X'\right), \tag{1.7}$$

or

$$\Pr\left(\mathbf{BR}_i(y^*_{-i};\beta_0) \ni y^*_i \,\middle|\, \mathbf{x} \in X'\right) > \Pr_\beta\left(\mathrm{BR}_i(y^*_{-i};\beta) = \{y^*_i\} \,\middle|\, \mathbf{x} \in X'\right). \tag{1.8}$$

Remark 1.6 Conditions (1.7) and (1.8) are formulated to accommodate the possibility that best response correspondences may not be single-valued. If $\mathbf{BR}_i$ was single-valued almost surely, then it would suffice to assume that:

$$\Pr\left(y^*_i \in \mathbf{BR}_i(y^*_{-i};\beta_0) \,\middle|\, \mathbf{x} \in X'\right) \neq \Pr\left(y^*_i \in \mathbf{BR}_i(y^*_{-i};\beta) \,\middle|\, \mathbf{x} \in X'\right), \tag{1.9}$$

as this would imply both conditions. We state the general formulation to accommodate models with robust indifference, such as reduced forms corresponding to

lute/minimax dominance introduced in Halpern and Pass (2012) and Salcedo (2012, §4).

[16] One of the drawbacks of our identification strategy is that it only allows to identify the marginal distributions of payoffs. With this approach, the joint distribution of payoffs can only be identified under either strong structural assumptions, or assuming conditional independence of payoffs across agents. See Section 1.4.2.

non-trivial extensive form games. In such cases, it is not sufficient to impose (1.9) because, even conditional on the event $\mathbf{y}_i \in \mathbf{BR}_i$, the probability of choosing y_i^* may be different from the probability that it is *one of* the best responses to y_{-i}^*.

Different standard sets of conditions imply Assumption 7. For example, one could assume independent action specific additive residuals with a known distribution as Bajari et al. (2010b), or semi-parametric single index models as Fox (2007). In any case, these assumptions are sufficient to point identify β_0 without imposing more restrictive solution concepts.

Proposition 1.4 *Under assumptions 1.1, 1.2, 6 and 7, β_0 is point identified.*

When best response correspondences are single-valued almost surely, Assumptions 6 and 7 imply Assumption 5, and therefore Proposition 1.4 follows from Proposition 1.3. To see this, simply take $Y_k = \{y_i^k\} \times Y_{-i}$, and $X_k^n = \{x_k \geq n\} \times X_{-k}$. However, this is not true in the general case, because h_0 may not be determined by rationality alone. The proof for the general case is in Appendix A.1.

1.4.2. Identification at infinity in the entry game

Let us proceed by establishing point identification of structural parameters in the entry game. Since vNM utility indexes are defined only up to positive affine transformations, we can normalize the variance of the error terms to 1 without loss of generality. The vector of unknown structural parameters is thus given by $\beta_0 = (\beta_{03}, \beta_{01}, \beta_{02}, \rho_0)^\mathsf{T} \in B \equiv \mathbb{R}_{++}^3 \times (-1, 1)$, where ρ_0 denotes the correlation between $\mathbf{e}_1$ and $\mathbf{e}_2$.

Note that, if $\mathbf{x}_i > (\beta_{03} + \mathbf{e}_i)/\beta_{0i}$, then entering is strictly dominant for player i. Similarly, if $\mathbf{x}_i < \mathbf{e}_i/\beta_{0i}$, then not entering is dominant. Since $\mathbf{e}_i$ is independent of $\mathbf{x}_i$, the probability of these events approaches 1 when $\mathbf{x}_i$ diverges to $+\infty$ or $-\infty$, respectively. The level-2 rationality assumption then implies that, as $\mathbf{x}_i$ approaches $+\infty$ $(-\infty)$ the probability that firm $-i$ chooses a best response to 1 (0) converges to 1. Since $\mathbf{u}_{-i}$ is independent of $\mathbf{x}_i$ conditional on $\mathbf{x}_{-i}$, Assumption 6 is satisfied.

In contrast, Assumption 7 need *not* be satisfied. The reason for this is that the best response correspondence for player i only depends on $\mathbf{e}_i$ and $\mathbf{x}_i$. Since the correlation parameter ρ_0 does not affect the marginal distribution of $\mathbf{e}_i$, it does not affect the distribution of best responses, and thus it cannot be identified using our strategy.

However, if we assume that ρ_0 is known, then the assumption is satisfied and the rest of the payoff parameters are point identified. Since best responses are generically single-valued, it suffices to verify that condition (1.9) holds. First note that:

$$\Pr\Big(1 \in \mathbf{BR}_i(1;\beta)\,\Big|\,x\Big) = \Pr\Big(\mathbf{e}_i < \beta_i x_i - \beta_3\Big) = \Phi\Big(\beta_i x_i - \beta_3\Big), \tag{1.10}$$

$$\text{and} \quad \Pr\Big(1 \in \mathbf{BR}_i(0;\beta)\,\Big|\,x\Big) = \Pr\Big(\mathbf{e}_i < \beta_i x_i\Big) = \Phi\Big(\beta_i x_i\Big), \tag{1.11}$$

where Φ denotes the standard normal c.d.f.. Given any $\beta \in B$, if $\beta_i \neq \beta_{0i}$, then (1.11) implies that (1.9) is satisfied for any set X' with at least two values. If $\beta_i = \beta_{0i}$ for both firms, but $\beta_3 \neq \beta_{03}$, then (1.10) implies that (1.9) is satisfied for any set X' with at least two values. This means that Assumption 7 is satisfied, and Proposition 1.4 thus implies that β_0 is point identified.

1.5. Identification of solution concepts

We are finally in a position to present our main result. Recall that a solution concept is consistent with the data if it is satisfied by some (β, h) belonging to the sharp identified set. In sections 1.3 and 1.4 we made assumptions guaranteeing that the sharp identified set essentially collapses to (β_0, h_0). Hence, under these assumptions, a solution concept is consistent with the data if and only if it is satisfied by the true structural parameters and distribution of play.

Theorem 1.5 *Under assumptions 1.1-5, a solution concept is consistent with the data if and only if it is satisfied by the players' behavior almost surely, i.e., if and only if* $\mathbf{h}_0 \in \mathbf{q}(\beta_0)$ a.s..

Proof. The result is a direct consequence of propositions 1.2 and 1.3. Suppose that q is consistent with the data, i.e., there exist (β, h) in the identified set which satisfy q. By proposition 1.3, we know that we have $\beta = \beta_0$. Since (β_0, h_0) always belongs to the identified set, it follows from proposition 1.2 that $\mathbf{h} = \mathbf{h}_0$ a.s.. Therefore, $\mathbf{h}_0 \in \mathbf{q}(\beta_0)$ a.s.. ■

Let us compare Theorem 1.5 with the sharp characterization in terms of Aumann expectations adapted from Beresteanu et al. (2011). Assuming that β_0 is point identified, Theorem 1.1 asserts that a solution concept is consistent with the data as long as $\mathbb{E}[\,\mathbf{h}_0|\mathbf{x}\,] \in \mathbb{E}[\,\mathbf{q}(\beta_0)|\mathbf{x}\,]$. The power of Theorem 1.5 is that, by point identifying h_0, we can dispense with these expectations and impose an ex-post condition, as if we could actually observe the distribution of choices conditional on both $\mathbf{x}$ and $\mathbf{e}$.

Remark 1.7 While Theorem 1.5 is correct as stated, the wording may be slightly misleading. To see this, consider the solution concept q_{NE} for the entry model. The definition of consistency requires the existence of some (β, h) in the identified set such that $\mathbf{h} \in \mathbf{q}_{\text{NE}}(\beta)$ a.s.. In terms of the underlying game, this requires that the distribution of play—integrating out any uncertainty arising from the selection mechanism—should belong to the set of NE. Because q_{NE} generically consists of locally isolated points, this implicitly assumes that firms always play the same equilibrium given $\mathbf{u}_0$. To remain agnostic about the selection criterion, one should consider convex solution concepts such as $\bar{q}_{\text{NE}}$. In other words, without strong assumptions about the way agents choose among equilibria, a solution concept can be identified only up to its convex closure.[17]

Remark 1.8 Admittedly, the assumptions used to guarantee point identification of β_0 are somewhat restrictive. However, the main insights of the result are independent of this fact. The exclusion restriction and the bounded completeness assumptions from section 1.3 point identify h_0 up to β_0, independently of whether β_0 is point or set identified.

1.5.1. Discriminating between competing solution concepts

Thus far, we have provided assumptions to identify whether a solution concept is consistent with the data, without specifying a particular alternative hypothesis. Now we consider the problem of discriminating between a given set of competing solution concepts under the assumption that at least one of them corresponds to the actual behavior of the agents.

Assumption 8 (β_0, h_0) jointly satisfy at least one "true" solution concept q_0 from a known set Q.

[17] This is true also in the sharp characterization from Beresteanu et al. (2011), since the conditional Aumann expectation of a convex set coincides with the conditional Aumann expectation of its convex closure. While we need the convex hull to integrate out uncertainty arising from selection, we still can treat $\mathbf{e}$ as observed and hence we do not require Aumann expectations.

Say that the solution concept is point identified if q_0 is the only solution concept in Q which is consistent with the data. By Theorem 1.5, q is consistent with the data only if $\mathbf{q}(\beta_0)$ contains the distribution of play almost surely. By assumption, we know that the distribution of play is in $\mathbf{q}_0(\beta_0)$ almost surely. Hence, as long as $\mathbf{q}(\beta_0)$ and $\mathbf{q}_0(\beta_0)$ are disjoint with positive probability, q cannot be consistent with the data. Point identification is thus attained whenever every pair of competing solution concepts make disjoint predictions with positive probability.

Theorem 1.6 *Under assumptions 1.1–5 and 8, if for every $q \in Q \setminus \{q_0\}$ there exists some $F \subseteq E \times X$ such that $\Pr((\mathbf{e}, \mathbf{x}) \in F) > 0$ and $q(\beta_0, e, x) \cap q_0(\beta_0, e, x) = \emptyset$ for almost all $(e, x) \in F$, then the solution concept is point identified.*

Proof. Let $q \in Q \setminus \{q_0\}$, and let F be such that $\Pr((\mathbf{e}, \mathbf{x}) \in F) > 0$ and $q(\beta_0, e, x) \cap q_0(\beta_0, e, x) = \emptyset$ for almost all $(e, x) \in F$. By assumption, we know that $\mathbf{h}_0 \in \mathbf{q}_0(\beta_0)$ a.s., and thus $\Pr\left(\mathbf{h}_0 \in \mathbf{q}(\beta_0) \,\middle|\, (\mathbf{e}, \mathbf{x}) \in F\right) = 0$. Hence, q is *not* satisfied by (β_0, h_0). Theorem 1.5 then implies that q is not consistent with the data. Since q was arbitrary, q_0 is the only solution concept in Q which is consistent with the data. ■

Remark 1.9 Even given assumptions 1.1–5 and 8, the condition from Theorem 1.6 is not necessary. Even if the intersection of q and q_0 is never empty, q could be ruled out if $\mathbf{h}_0 \in \mathbf{q}_0(\beta_0) \setminus \mathbf{q}(\beta_0)$ with positive probability. However, in terms of the underlying game, we believe that it is the weakest condition that is sufficient without imposing any assumptions of the equilibrium selection mechanism.

Remark 1.10 Note that the condition from Theorem 1.6 cannot help to discriminate between nested solution concepts. Indeed if $q \subseteq q'$ pointwise, then $q(\beta_0, e, x) \cap q_0(\beta_0, e, x)$ is never empty. If the smaller solution concept is satisfied, then so is the larger one. However, it may still be possible to rule out the difference $q' \setminus q$. Furthermore, it may be possible to rule out the smaller solution concept, but not without further assumptions on equilibrium selection. See, for instance section 1.6.2.

1.6. Applications

We have developed a general framework to identify which solution concepts are satisfied by actual behavior patters. In this section, we consider different specific examples to suggest potential applications, and illustrate the power and limitations of our methodology. First, we show that it is possible to identify whether entry choices satisfy equilibrium conditions. Then, we illustrate one way to deal with nested solution concepts, by identifying whether firms in the entry game play mixed strategies. In the third application, we identify the likelihood of different equilibria in an n-player coordination game. Then, we distinguish between simultaneous and sequential choices in the entry game, to illustrate how to handle extensive form games. Finally, we consider the problem of discriminating between complete and private information in the entry game.

1.6.1. Are firms' choices in equilibrium?

Many theoretical results and empirical exercises rely on the idea that firm behavior must converge to equilibrium. Estimation is often carried out under the assumption that observed data arises from equilibrium choices, and policy implications are derived from counterfactual analyses which compares equilibria under alternative policies given the estimated parameters. Our results allow to identify the validity of this approach, at least in our simple entry environment, assuming level-2 rationality and independent error terms, and assuming that the distribution of play is independent of $\mathbf{x}$ conditional on $\mathbf{u}_0$.

In sections 1.3 and 1.4 we have shown that these assumptions imply that the assumptions from Theorem 1.5 are satisfied. Therefore, there exists a selection mechanism and payoff parameter that rationalize q_{NE}, if and only if the firms choices are in equilibrium almost surely. Hence, this opens up the possibility of testing whether and how often firm behavior actually satisfies equilibrium conditions.

1.6.2. Do firms play mixed strategies?

Assuming Nash play in our entry model, is it possible to identify whether firms play mixed strategies or not?[18] Let $\bar{q}_{\mathrm{PNE}}$ be the solution concept that restricts firms

[18] This question has been addressed in different contexts, for instance, by Chiappori et al. (2002) and Palacios-Huerta (2003). They analyze data from penalty kicks in professional football to investigate whether it resembles a mixed strategy equilibrium of a zero-sum game. There are two important differences between their approach ours. First, they assume that the game is zero-sum, which eliminates the possibility of multiple equilibria. Second, instead of using a structural model, they test two reduced form implications of Nash play: indifference between alternatives, and serial independence.

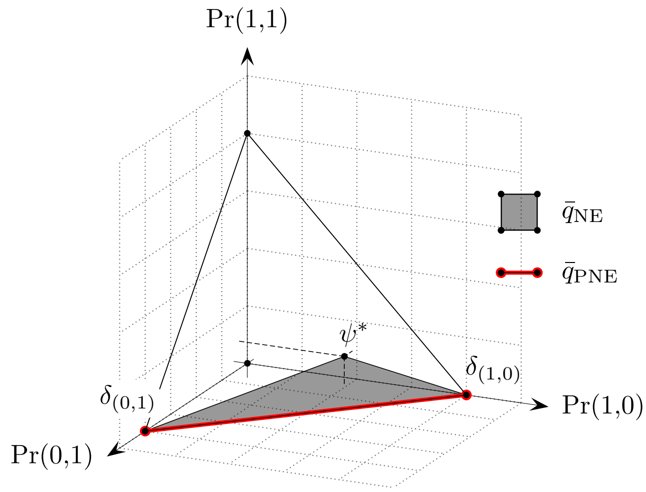


Figure 1.4 – Predictions for pure vs. mixed NE in the multiplicity region for the entry game. The point ψ^* corresponds to the mixed NE.

to playing NE in pure strategies. Theorem 1.6 is not directly applicable because $\bar{q}_{\mathrm{PNE}} \subseteq \bar{q}_{\mathrm{NE}}$. In particular, if $\bar{q}_{\mathrm{PNE}}$ is consistent with the data, then so is $\bar{q}_{\mathrm{NE}}$. However, it is indeed possible to determine whether firms play mixed strategies.

For this purpose, it suffices to assume that the error terms are independent, so that the payoff parameters are identified at infinity, see section 1.4.2. In the multiplicity region, all pure equilibria correspond to monopolies. Therefore, under $\bar{q}_{\mathrm{PNE}}$, a duopoly arises only when it is strictly dominant for both firms to enter the market, and the probability of this event can be computed knowing the value of the structural parameters. In contrast, because the randomization from mixed strategies is independent across firms, every mixed equilibrium assigns positive probability to the outcome $(1, 1)$. Therefore, if firms play mixed strategies with positive probability, then the probability of observing a duopoly is strictly larger than under $\bar{q}_{\mathrm{PNE}}$.

There are two reasons why we can identify whether firms play mixed strategies or not, despite having nested solution concepts. First, in order to rule out $\bar{q}_{\mathrm{PNE}}$, we make a mild assumption about the selection criteria, namely, that the mixed equilibrium is chosen with positive probability. Second, when we want to rue out mixed strategies, we do not attempt to rule out $\bar{q}_{\mathrm{NE}}$ (which is assumed to hold). Instead, we *locally* rule out the solution concept $\bar{q}_{\mathrm{NE}} \setminus \bar{q}_{\mathrm{PNE}}$ according to which firms *sometimes* play mixed strategies. This is possible in this example because the particular structure of the equilibrium correspondence guarantees that, in the multiplicity region, the pure equilibria lie on an edge of the simplex while the mixed equilibria always lies on the interior, see Figure 1.4.

Note that for this application we did not require point identification of the distribution of play, and hence we did not use the exclusion restriction nor the completeness of the payoff distribution. With this additional structure, we could actually determine exactly how often do firms play mixed NE, and how does the probability of choosing each equilibrium changes with the realization of the payoff functions. We now proceed to discusses an application in which these questions are more relevant.

1.6.3. Equilibrium selection in coordination games

The role of selection criteria is specially relevant in coordination games with multiple equilibria that are ranked in the Pareto sense. Different theoretical literatures suggest that agents are likely to coordinate on risk-dominant rather than efficient equilibria (e.g., Harsanyi and Selten (1988) and Carlsson and Van Damme (1993)). This has important welfare implications as it opens the possibility of welfare-improving policies. However, experimental evidence suggests that agents may coordinate on different equilibria depending on the specific payoffs (Battalio et al., 2001). Our methodology might bring some light to this issue by identifying the likelihood of different equilibria for different real life situations, and how these likelihoods depend on the characteristics of the environment.

To fix ideas, we analyze a particular n-player coordination game modelling a regime-change environment. However, similar payoff structures can be used to model various situations including coordinated attack problems (Rubinstein, 1989), bank-runs and currency attacks (Morris and Shin, 2003), or tacit collusion in oligopolistic markets Green et al. (2013).

Consider a small village with citizens $i \in I = \{1, 2, \ldots, \iota\}$. At a given time, each citizen chooses whether to manifest discontent towards the current regime (revolt) or not, $y_i \in \{0, 1\}$. The regime is changed if and only if the proportion of citizens revolting is greater than some threshold given by

$$\mathbf{t} = \Phi(\eta_0^{\mathsf{T}}\mathbf{w} + \eta_{0I}\mathbf{z}_I + \mathbf{e}_I),$$

where Φ is the standard normal p.d.f.. We normalize the payoff from not revolting to 0, and assume that the payoff from revolting is given by

$$\mathbf{u}_{0i}(1, y_{-i}) = \begin{cases} \lambda_0^{\mathsf{T}}\mathbf{w} + \lambda_{0i}\mathbf{z}_i + \lambda_{0i}^*\mathbf{z}_i^* + \mathbf{e}_i^* & \text{if} \quad \sum_i y_i > \mathbf{t} \cdot \iota \\ \gamma_0^{\mathsf{T}}\mathbf{w} + \gamma_{0i}\mathbf{z}_i + \gamma_{0i}^*\mathbf{z}_i^* + \mathbf{e}_i & \text{otherwise} \end{cases}.$$

In our specification $\mathbf{z} = (\mathbf{z}_I, (\mathbf{z}_i, \mathbf{z}_i^*)_{i\in I})$ is the vector of excluded covariates satisfying Assumption 1.3, $\mathbf{x} = (\mathbf{w}, \mathbf{z})$ is the vector of observed covariates, $\mathbf{e} = (\mathbf{e}_I, (\mathbf{e}_i, \mathbf{e}_i^*)_{i\in I})$ is the vector of error terms, and $\beta_0 = (\eta_0, \lambda_0, \gamma_0, (\lambda_{0i}, \lambda_{0i}^*, \gamma_{0i}, \gamma_{0i}^*)_{i\in I})$ is the vector of unknown structural parameters.

We maintain the assumption that payoffs are common knowledge and the citizens play NE, but we remain agnostic about which equilibria are more likely to arise. A coordination problem arises when citizens want to manifest discontent if and only if the revolt is successful, and $\mathbf{t} > 1/\iota$, meaning that more than one player is required for a successful revolt. In that case, there is a NE in which no one revolts, and it is Pareto dominated by a different NE in which everyone revolts. However, depending on the realized utility functions, the inefficient NE might be risk-dominant.

We do not require a lot of structure on $\mathbf{w}$ nor the signs of the parameters. We simply assume that (i) the distribution of the error terms belongs to the exponential family, (ii) the excluded covariates $\mathbf{z}$ are continuous and their coefficients are different from zero, and (iii) the support of $\mathbf{z}_I$ and $\mathbf{z}_i$ for $i \in I$ is the entire real line.[19] Covariate $\mathbf{z}_I$ should be something that affects the strength or resiliency of the current regime. For instance, it could be an indicator of the financial health of the regime. Each $\mathbf{z}_i$ and $\mathbf{z}_i^*$ should affect how much citizen i cares about changing the regime or about being on good terms with the current regime (in case the revolt fails). For instance, one could use household income or, better yet, an proxy for the proportion of i's business that depend on the current regime.

Our identification-at-infinity approach can be used in this setting to point identify β_0. However, even if β_0 was not point identified, our assumptions guarantee that Proposition 1.2 applies. Therefore h_0 is point identified, at least up to β_0. This makes it possible to recover from the data the average distribution of choices for each possible realization of the vNM indexes. Hence, we can measure the exact welfare loss arising from miss-coordination or coordination on inefficient equilibria, and the probability that each kind of equilibrium is selected as a function of the characteristics of the environment.

[19] We use the same covariates to identify payoff parameters and the distribution of play, but we do so only to keep the notation simple. In general, we do need some excluded covariates and some unbounded covariates, but they need not be the same.

1.6.4. Identifying the sequential structure of choices

In the previous examples we assumed that the game was a simultaneous move game. However, our framework can easily accommodate more general environments. In this section, suppose that the structure of the game is unknown to the econometrician. In particular, it is known that the agents play subgame perfect Nash equilibria (SPNE), but it is not known whether the firms choose their actions simultaneously or sequentially, or which firm moves first. To keep things simple, we consider three possible extensive form games corresponding to three possible game trees, two of which are illustrated in Figure 1.5. Let $\bar{q}_i$ be the solution concept restricting behavior to SPNE of the the game in which firm i moves first.

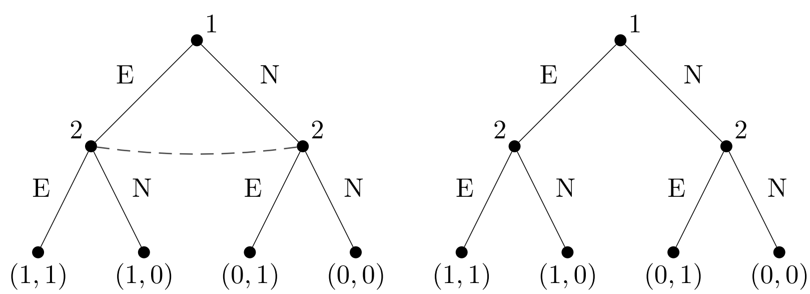


Figure 1.5 – Simultaneous move entry game (left panel) and sequential entry game with player 1 moving first (right panel).

It is straightforward to verify that our previous arguments to establish identification of β_0 and h_0 still apply. Hence, in order to identify point identification of $q_0 \in \{\bar{q}_{\text{NE}}, \bar{q}_1, \bar{q}_2\}$, it suffices show that the convex hull of the predictions of our competing models are disjoint for some realizations of $\mathbf{x}$ and $\mathbf{e}$. To see that $\bar{q}_1$ is distinguishable from $\bar{q}_2$, recall that in the multiplicity region, a firm wants to enter if and only if his opponent does not enter. That means that the game in which i moves first has a unique SPNE outcome, in which i is the only entrant. Hence, q_1 and q_2 make disjoint predictions in this region.

In contrast, without restrictions on selection, $\bar{q}_{\text{NE}}$ cannot be distinguished from $\bar{q}_i$, because the structure of payoffs guarantees that $\bar{q}_i \subseteq \bar{q}_{\text{NE}}$ for all realizations of payoffs. This point has been raised, for instance, by (Bresnahan and Reiss, 1990, §3.3). We overcome this issue by assuming that there is a different covariate $\mathbf{z}_1$ which takes both positive and negative values with positive probability and such that:

$$\mathbf{u}_1(y) = (\beta_{01}\mathbf{x}_1 - \beta_{03}\mathbf{z}_1 y_2 - \mathbf{e}_1)y_1.$$

That is, firm 1 sometimes benefits from having firm 2 in the market. This assumption can be justified considering asymmetric retailers. Suppose that firm 2 is a large departmental store with a well renown brand, and firm 1 is a small local firm. Firm 1 may benefit from having firm 2 nearby, as firm 2 may attract a large costumer flow, while firm 2 may still prefer to be a monopolist.

With this new covariate, it suffices to consider the set:

$$\left\{(x, e, z_1) \in \mathbb{R}^5 \,\middle|\, 0 < \beta_2 x_2 - e_2 < \beta_3 \quad \wedge \quad \beta_3 z_1 < \beta_1 x_1 - e_1 < 0\right\}.$$

In this region, the simultaneous move game only admits an asymmetric mixed strategy equilibrium, while the sequential games only admit pure strategy equilibria, and we thus have $\bar{q}_{\text{NE}} \cap \bar{q}_i = \emptyset$. Therefore, by Theorem 1.6, all the equilibrium concepts are distinguishable, and q_0 is point identified.

1.6.5. Incomplete information games

For our final example, we discriminate between two informational assumptions in our entry model, assuming simultaneous choices and equilibrium condition.[20] In particular, we contrast the assumption that payoffs are common knowledge, versus the alternative hypothesis that each firm i only observes $\mathbf{x}$ and $\mathbf{e}_i$, but not $\mathbf{e}_{-i}$. Our general results are not directly applicable to incomplete-information games, however, for this particular example, we can still identify the information structure.

Under the incomplete-information hypothesis, a pure strategy for i is a function $s_i : E_i \to \{0, 1\}$. The game only admits pure strategy BNE, and a strategy profile s^* is a BNE if and only if it can be written as:

$$s_i^*(e_i) = \begin{cases} 0 & \text{if} \quad e_i > \bar{e}_i(x) \\ 1 & \text{if} \quad e_i < \bar{e}_i(x) \end{cases}, \tag{1.12}$$

where $\bar{e}(x) = (\bar{e}_1, \bar{e}_2)$ satisfies:

$$\bar{e}_i(x) = \beta_{0i} x_i - \beta_{03} \Phi(\bar{e}_{-i}(x)), \tag{1.13}$$

[20]This application is closely related to the work of Grieco (2014), who constructs estimates for structural parameters in an entry environment that are robust to different informational assumptions. In contrast, we do not focus on the robust identification of structural parameters, but rather on the possibility to discern between different information structures.

for $i = 1, 2$. There always exists at least one such equilibrium, and there are multiple equilibria for some values of x, see Appendix A.2 for more details. Let $\bar{q}_{\text{BNE}}$ be the solution concept corresponding to BNE of the incomplete-information game.

The identification at-infinity-approach applies under both solution concepts, and thus β_0 is point identified as long as the error terms are independent across firms. The difficulty arising from incomplete information regards the distribution of play. The problem is that the set of choices that are consistent with BNE depends not only on the realized utility functions, but also on the value of the covariates. To see this, let (x, e) and (x', e') be such that $x'_1 > x_1$, and:

$$e'_i = e_i + \frac{x_1 - x'_1}{\beta_{10}}.$$

Then, the realized vNM indexes functions are the same, but $\bar{e}(x)$ and $\bar{e}(x')$ differ. This implies that $\bar{\mathbf{q}}_{\text{BNE}}(\beta)$ is not measurable with respect to $\mathbf{u}_\beta$, and, if $\bar{q}_{\text{BNE}}$ were the true solution concept, then h_0 would *not* satisfy Assumption 1.3.

In this particular game, it is still possible to distinguish $\bar{q}_{\text{NE}}$ from $\bar{q}_{\text{BNE}}$, but we have to rely solely on Theorem 1.1. That is, even if we cannot identify the distribtuion of choices conditional on $(\mathbf{e}, \mathbf{x})$, we can still discriminate between the solution concepts if the set of their average predictions conditional only on $\mathbf{x}$ are disjoint. Since the structural parameter is still point identified, it suffices to find a set of covariate values $X' \subseteq X$ realized with positive probability, and such that:

$$\mathbb{E}\left[\,\bar{\mathbf{q}}_{\text{NE}}(\beta_0)\middle|X'\,\right] \cap \mathbb{E}\left[\,\bar{\mathbf{q}}_{\text{BNE}}(\beta_0)\middle|X'\,\right] = \emptyset.$$

In Aappendix A.2, we show that, as x diverges to infinity along a specific direction, the maximum probability of a duopoly under any NE of the complete information game is strictly less than the minimum probability under any BNE of the incomplete information game. Hence, far enough along this direction, the corresponding Aumann expectations are strictly separated by a hyperplane—orthogonal to the degenerate distribution which assigns full probability to a duopoly—and are thus disjoint, see Figure 1.6.

Proposition 1.7 *Let $x(t) \in X$ be given by $x_i(t) = \beta_{03}t/\beta_{0i}$. There exists some $t_0 \in \mathbb{R}$ such that $M_{\text{NE}}(t) < m_{\text{BNE}}(t)$ for every $t \geq t_0$, where:*

$$M_{\text{NE}}(t) \equiv \max\left\{\psi(1,1)\,\middle|\,\psi \in \mathbb{E}[\,\bar{\mathbf{q}}_{\text{NE}}(\beta_0)|x(t)\,]\right\},$$

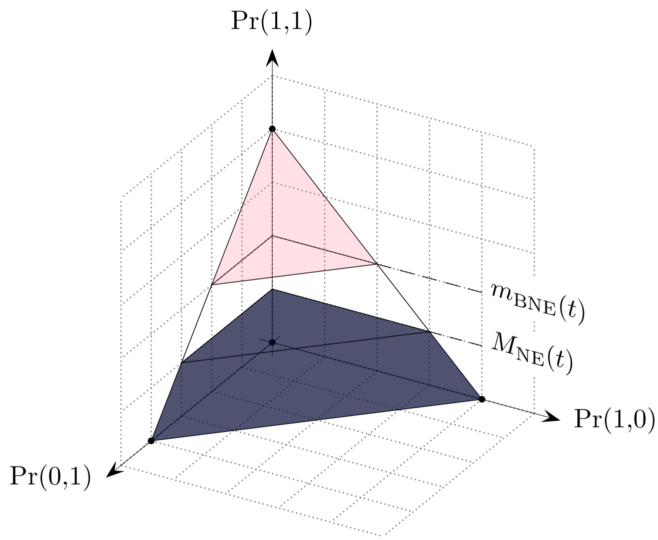


Figure 1.6 – Separation of $\mathbb{E}[\,\bar{\mathbf{q}}_{\mathrm{NE}}(\beta_0)|x(t)\,]$ (contained in the light shaded region) and $\mathbb{E}[\,\bar{\mathbf{q}}_{\mathrm{BNE}}(\beta_0)|x(t)\,]$ (contained in the dark shaded region) in the entry game, for sufficiently high values of t.

$$m_{\mathrm{BNE}}(t) \equiv \min\left\{\psi(1,1)\,\middle|\,\psi \in \mathbb{E}[\,\bar{\mathbf{q}}_{\mathrm{BNE}}(\beta_0)|x(t)\,]\right\}.$$

Remark 1.11 The sets $X(t_0) \equiv \{x(t) \mid t \in [t_0,\infty)\}$ are null, and we need a set X' with positive probability. This is not an issue because the separation is strict, and the corresponding Aumann expectations are continuous in x, therefore a set with positive measure can be obtained by considering small open neighbourhoods of $X(t_0)$.

While the equilibrium concepts can be identified in this case, the derivation is far from trivial even in this simple example. In particular, it requires a full characterization of the set of equilibria for all possible realizations of the residuals, and integrating over a selection of extreme points of these sets. This helps to illustrate the value of Proposition 1.2, and suggests the need to further investigate a more practical approach for models with incomplete information.

1.7. Conclusion

The current work provides a framework to identify the behavior of agents conditional on both observed and unobserved heterogenity. This opens up the possibility to test a number of important assumptions regarding behavior patterns, solution concepts, and structural features (such as the timing of choices), using real life data. For example, it makes it possible to test whether firm's entry choices satisfy equilibrium conditions, and which equilibria are more likely to arise in coordination problems. Testing these features is crucial, not just to estimate structural parameters, but also to verify the validity of counterfactual analyses and policy implications.

Our results apply to a general class of discrete complete-information games both in strategic and extensive form. In its current state, our general methodology cannot be directly applied to incomplete-information games, and relies on semi-parametric assumptions. It might be possible to deal with incomplete information if the researcher observes covariates that were private information at the moment the game was played, but became public after choices where made. This kind of data is not uncommon. For instance, the financial characteristics of a firm may only become public at the end of the financial year. As for the parametric restrictions, it is hard to conceive a full non-parametric approach, but it may be feasible to allow for much greater flexibility by introducing random coefficients. We leave these as open problems for future research.

Chapter 2

Testing Nash assumption

Abstract *Nash equilibrium* (NE) is a leading solution concept in the empirical analysis of entry games. The NE assumption is crucial, not just for estimation, but also for the validity of counterfactual exercises and policy implications. I propose a computationally simple sieve likelihood ratio type procedure to test the NE assumption in a complete information entry game with *second-order rational* players. The method is robust to partial identification and allows for nonparametric selection of equilibria.

The empirical literature on parametric complete information entry games, started by Berry (1990) and Bresnahan and Reiss (1990), has focused on the estimation of payoff parameters under the assumption of *Nash equilibrium* (NE) behavior of the players.[1] For instance, Ciliberto and Tamer (2009) analyze the market structure of the US airline industry assuming NE play in pure strategies.[2] The importance of testing the NE assumption goes beyond estimation. Even if the payoff parameters are known or can be consistently estimated a false NE assumption can lead to dramatically different policy implications, see Kashaev and Salcedo (2016). The validity of the NE assumption has not been fully addressed in settings where payoffs are unknown.[3]

[1]Relevant examples include papers about discrete complete-information games Bjorn and Vuong (1984), Tamer (2003), Aradillas-López and Tamer (2008), Bajari et al. (2010b), Beresteanu et al. (2011), Galichon and Henry (2011), Henry and Mourifie (2012), Kline and Tamer (2012), Aradillas-López and Rosen (2013), Bresnahan and Reiss (1991), Berry and Tamer (2006), Ciliberto and Tamer (2009), Kline (2015a).

[2]Their procedure can be extended to the case where players are allowed to play mixed strategy NE.

[3]Similar empirical questions have been studied in the experimental game theory literature; e.g., Camerer (2003), Crawford et al. (2013), Kline (2015b).

This paper aims to fill this gap.

The objective of this paper is to test whether the NE assumption is consistent with the observed data on payoff shifters $\mathbf{x}$ and outcomes $\mathbf{y}$ in semiparametric entry games with complete information. I illustrate my methodology by considering a two player entry game with *second-order rational* (SOR) agents. That is, I assume that every player is a profit maximizer and knows that her opponent is maximizing her payoff as well.

First, under the SOR assumption, I describe the distribution of outcomes conditional on covariates, $p(\mathbf{y}|\mathbf{x};\theta)$, as a function of $\theta = (\beta, h)$, where β is a finite dimensional parameter governing the distribution of payoffs and h, *the distribution of play*, is an infinite dimensional parameter describing behavior of players. Second, I show that the NE assumption can be characterized by a finite set of equality and inequality constraints on θ. Third, I propose a sieve likelihood ratio (LR) type testing procedure by taking the difference between properly normalized unconstrained (implied by SOR) and constrained (implied by NE) maximum likelihood objective functions. The procedure does not assume point identification of parameters, does not impose parametric assumptions on the way players randomize among different equilibria in the region of multiplicity of NE, and can be extended to entry games with more than two players.

The SOR assumption is essential for the analysis. It is general enough to allow agents to play any strategy that survives two rounds of strict dominance elimination, e.g., Nash or correlated equilibrium. Moreover, the players can be *ambiguity averse* when they are not sure what the opponent is going to play. However, the agents are not allowed to play maximin strategies under the SOR assumption. For a detailed discussion see Aradillas-López and Tamer (2008).

My approach to modeling entry decisions is different from the existing literature in several respects. First, the standard approach is to assume a particular equilibrium concept, say NE, and then augment the model with a selection mechanism.[4] Since the selection mechanism is defined up to an equilibrium concept, every choice of equilibrium concept results in a different model. In contrast, the distribution of play, h, I am using, is the probability that a particular *outcome (not equilibrium)* is played conditional on a realization of covariates and unobservables. Then an equilibrium concept assumption is just a set of restrictions on parameters. So, I parametrize the model independently of the equilibrium concept. Second, I do not treat h as a

[4]For instance, Bajari et al. (2010b) uses a parametric selection mechanism and Beresteanu et al. (2011) uses a nonparametric one.

nuisance parameter. The distribution of play itself is an important object. It can be informative about the way agents mix among different outcomes when NE does not give point predictions. For instance, as an intermediate step of my procedure, one can test whether agents play NE in pure strategies only or play the mixed strategy NE as well.

Although I am focusing on testing the NE assumption, the way I parametrize the model allows me to test any behavioral restrictions as long as these restrictions are consistent with SOR and can be characterized by a set of equality/inequality constraints that are smooth in β and affine in h.

I work with a likelihood based model for two reasons. First, in the point identified case it is efficient. Second, many conditional moment equalities/inequalities models do not use all available information, in which case, the regions of parameters they propose are not sharp.[5]

I test whether one equality and three inequality constraints are satisfied for some point in the identified set. Unfortunately, one cannot know *a priori* which inequality constraints are binding. I could have focused on the least favorable null by replacing inequality constraints by equality constraints. However, since at most two inequality constraints can be binding simultaneously, it is not clear what the least favorable null is. Moreover, such a procedure is extremely conservative. To overcome this problem I propose to use a sequence of statistics to test whether any combination of the inequality constraints is binding. In particular, I start with testing whether only two of the inequality constraints are binding. Next, I test whether there is only one binding inequality constraint. Finally, I consider the case in which none of the inequality constraints is binding.

Deriving the asymptotic null distribution of the statistics creates some important challenges. The NE assumption imposes restrictions on the behavior of players that have to be satisfied at an infinite number of points. That is, there are infinitely many NE constraints. I overcome this problem by aggregating infinite many constraints into finitely many ones. Next, I show that the finite dimensional part of the constrained sieve maximum likelihood estimator converges to the identified set in one sided Hausdorff distance at an appropriate rate. This result allows me to linearize the NE constraints. Finally, I show that the asymptotic null distribution of the statistics is defined by the empirical process indexed by Riesz representers of the linearized

[5]The identified set implied by the likelihood model is always a subset of the region of parameters consistent with the system of conditional moment equalities/inequalities. See Section 3.3 in Beresteanu et al. (2011).

NE constraints.[6]

An alternative approach for testing the NE assumption is to check whether the confidence set for the payoff parameters under the NE assumption is empty. Chen et al. (2011) (CTT) use the profiled LR statistic to build a confidence set which is never empty by construction. Andrews et al. (2004), Ciliberto and Tamer (2009), Aradillas-López and Rosen (2013), Epstein et al. (2015), among others, have developed techniques to construct confidence sets for the payoff parameters in partially identified games either by ruling out mixed strategy NE, or by considering a system of conditional moment equalities/inequalities *implied* by the NE assumption. However, as mentioned before, if one uses a confidence set based on a system of conditional moment equalities/inequalities, as in Ciliberto and Tamer (2009), then this confidence set can be nonempty with probability approaching one even if the NE assumption is violated.

To the best of my knowledge, the only papers that work with sharp identified sets without ruling out mixed strategy NE in the conditional moment equalities/inequalities setting are Beresteanu et al. (2011) and Galichon and Henry (2011). Their methods are based on support functions of the convex sets that are predicted by the model, and on the core of Choquet capacity respectively. Once a system of moment equalities/inequalities is built they propose to use existing methods to check whether or not the confidence set is empty. Since my procedure is based on a different criterion it is hard to compare my test to theirs directly. However, *(i)* testing the NE assumption using the system of moment equalities/inequalities is a *by-product* of the construction of confidence sets for the payoff parameters; *(ii)* because of the likelihood nature of my procedure, I expect it to perform better at least in the point identified case; *(iii)* under my setting it is easy to impose additional restrictions, such as exclusion and/or monotonicity restrictions, on h. It is worth noting that the methodology developed in Chernozhukov et al. (2015) can be applied to my problem also.

The paper is closely related to the literatures on (semi)parametric LR or QLR statistics. For instance, Murphy and Van der Vaart (2000) and Shen and Shi (2005) establish the chi-squared limiting distribution of the profiled LR and sieve LR statistics when parameters are point identified and regular. Chen and Liao (2014) derive similar results for point identified and irregular parameters. Liu and Shao (2003) derive an asymptotic null distribution of the parametric LR statistic when the parameters are partially identified. CTT extends Liu and Shao (2003) and Shen and

[6] Similar results in different settings for the semiparametric LR, quasi-likelihood ratio and sup-QLR statistics were obtained in Chen et al. (2011) and Tao (2014).

Shi (2005) to the case with set identified infinite dimensional parameters. Chen and Pouzo (2015) provide the limiting distribution of the QLR statistic for point identified and possibly irregular parameters. Tao (2014) derives the limiting distributions of the QLR and sup-QLR statistics for the class of partially identified models. The main results in CTT and Tao (2014), despite their similarity to my approach, cannot be applied to my problem.

Since the NE assumption is a model assumption, the paper also contributes to the literature on model specification testing. In completely parametric or point identified settings numerous results were obtained; e.g., Vuong (1989), Kitamura (2000), Rivers and Vuong (2002), Hsu and Shi (2013). My paper is an example of a semiparametric model specification test of nested models under partial identification.

Once the NE assumption is tested, one can employ existing methods to construct confidence sets for parameters of the model, as in CTT and Ciliberto and Tamer (2009).

My testing procedure is based on the finite-dimensional sieve maximum likelihood approach. Hence, it inherits all its advantages. Once the infinite dimensional parameter is replaced with its sieve approximation, for implementation purposes the problem becomes parametric and is analogous to parametric MLE. So, it is computationally simple.

The paper is organized as follows. Section 2.1 describes the setting of the game and assumptions on observables and unobservables. Section 2.2 describes the constrains imposed on the model by the SOR and NE assumptions in the two player entry game. In Section 2.3 I describe assumptions on the parameter space and define the identified set. Section 2.4 provides results on consistency and the rate of convergence of the sieve maximum likelihood estimator in terms of the Pearson distance. Section 2.5 describes the testing procedure and the null distribution of statistics involved. I provide results on the multiplier bootstrap in Section 2.6. In Section 2.7 I provide results for Monte Carlo experiments and empirical application based on the model and data presented in Grieco (2014). Section 2.8 concludes.

2.0.1. Notation and definitions

I use bold font for random objects and regular font for their realizations. For a column vector $\beta \in \mathbb{R}^{d_\beta}$, β^T, β_i, β_{-i}, and $\|\beta\|_e$ denote its transpose, the i-th component, $(\beta_1, \dots, \beta_{i-1}, \beta_{i+1}, \dots, \beta_{d_\beta})$, and its Euclidean norm respectively; $\|\cdot\|_\infty$ denotes the sup-norm. For $f : Y \to \mathbb{R}$, where $Y = \{y_1, y_2, \ldots, y_k\}$ is a finite set, $(f(y))_{y \in Y}$ denotes $(f(y_1), f(y_2), \ldots, f(y_k))^\mathsf{T}$. For $f : X \times \Theta \to \mathbb{R}$, where $X \subset \mathbb{R}^{d_x}$, $\partial_x f(x_0, \theta_0)$ denotes the vector of partial derivatives of $f(\cdot, \theta_0)$ evaluated at x_0; $f(\theta_0) = f(\cdot, \theta_0)$ and $\partial_{x^\mathsf{T}} f(x_0, \theta_0) = (\partial_x f(x_0, \theta_0))^\mathsf{T}$. I use f_{x_1} and $f_{x_1|x_2}$ to denote the probability density functions (p.d.f.) of a continuously distributed $\mathbf{x}_1$ and of $\mathbf{x}_1$ given $\mathbf{x}_2 = x_2$. For $d \in \mathbb{R}$, $[d]$ denotes the integer part of d. For $\{\mathbf{x}_n\}_{n=1}^{\infty}$ and $\{\alpha_n\}_{n=1}^{\infty}$, $\mathbf{x}_n = O_p(\alpha_n)$ and $\mathbf{x}_n = o_p(\alpha_n)$ mean that $\{\mathbf{x}_n/\alpha_n\}_{n=1}^{\infty}$ is bounded in probability and converges in probability to 0 respectively.

2.1. An entry game

2.1.1. Payoff functions

Consider a two player entry game with complete information. Two firms $i \in I = \{1, 2\}$ must decide whether to enter a market ($\mathbf{y}_i = 1$) or not ($\mathbf{y}_i = 0$). I do not specify the strategy space for players. Instead, let $Y = \{(0,0), (1,1), (1,0), (0,1)\}$ be the set of outcomes of the game. The payoff of player i for an outcome $y \in Y$ is given by:

$$u_i(y, x, e, \beta) = (\bar{x}_i^\mathsf{T} \bar{\beta}_i - \tilde{x}_i^\mathsf{T} \tilde{\beta}_i y_{-i} - e_i) y_i, \tag{2.1}$$

where $\mathbf{x}$ is the observed vector of covariates that is comprised of non-redundant elements of $\bar{x}_i$, $\tilde{x}_i$, $i = 1, 2$. Let $X \subseteq \mathbb{R}^{d_x}$ be the support of $\mathbf{x}$. $\mathbf{x}$ is observed by both players and the econometrician; the part of payoffs unobserved by econometrician is $\mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2)^\mathsf{T}$. I will assume that the distribution of $\mathbf{e}|(\mathbf{x} = x)$ is known to the econometrician up to a finite dimensional parameter. Payoff shifters $\bar{x}_1$, $\bar{x}_2$, $\tilde{x}_1$ and $\tilde{x}_2$ can have common components.

The term $\bar{x}_i^\mathsf{T} \bar{\beta}_i - e_i$ represents i's benefit or cost of entering the market, while $\tilde{x}_i^\mathsf{T} \tilde{\beta}_i$ represents i's cost of competition when both firms enter. Let $\beta \in B \subseteq \mathbb{R}^{d_\beta}$ be a finite dimensional parameter vector governing the distribution of payoffs. It consists of $\bar{\beta}_i$, $\tilde{\beta}_i$, $i = 1, 2$, and the parameters of the conditional distribution of unobservables.

I assume the linear specification of payoffs for the sake of exposition. All the assumptions can be reformulated for a general parametric specification of payoff distribution as long as the payoffs are additive separable in e and sufficiently smooth and x and β.

Assumption 1 Covariates:

(i) $\mathbf{x}$ is continuously distributed with p.d.f. f_x which is bounded and bounded away from zero on the interior of X;

(ii) $X = \times_{i=1}^{d_x} X_i$, where X_i is a compact interval in $\mathbb{R}$;

(iii) $\mathbb{E}\left[\mathbf{z}\mathbf{z}^{\mathsf{T}}\right]$ is a positive definite matrix for $z \in \{\bar{x}_1, \bar{x}_2, \tilde{x}_1, \tilde{x}_2\}$;

Unobservables:

(i) $\mathbf{e}|\mathbf{x} = x$ has support $E = \mathbb{R}^2$ for all $x \in X$ and admits a conditional density function $f_{e|x}$ that is known up to a finite dimensional parameter that is a part of β, and is strictly positive on the interior of its support;

(ii) $f_{e|x}$ is continuous and bounded on $E \times X \times B$; $\int_E \left\|\partial_\beta f_{e|x}(e|x;\beta)\right\|_e de$ is continuous on $X \times B$;

(iii) $f_{e|x}$ is $[d_x/2] + 1$-times continuously differentiable in x for all $\beta \in B$;

Assumptions 1.(i) and 1.(ii) are not very restrictive. I can accommodate discrete covariates by introducing sieve spaces for each point in their support. Assumption 1.(iii) is a rank condition for single index models. Assumption 1.(i) is standard in the literature and guarantees that every outcome realizes with positive probability conditional on the realization of covariates. Assumptions 1.(ii) and 1.(iii) are satisfied by many parametric distributions. For instance, one can assume that $(\mathbf{e}_1, \mathbf{e}_2)$ are independent of $\mathbf{x}$; jointly normal with unknown correlation parameter ρ.[7]

Assumption 2 $\Pr(\inf_{\beta \in B} \tilde{\mathbf{x}}_i^{\mathsf{T}} \tilde{\beta}_i > 0) = 1$, $i = 1, 2$;

Assumption 2 is standard in the literature and requires firms to prefer monopolies over duopolies.

2.1.2. Distribution of play

The above model is incomplete. I have neither specified the order of firms' actions (simultaneous or sequential) nor the behavioral assumptions. Next, I describe the

[7] For normally distributed $\mathbf{e}$ variances are usually normalized to be 1. One can set $\bar{\beta}_{1,i} = 1$, $i = 1, 2$, instead.

behavior of players using the notion of *the distribution of play* introduced in Kashaev and Salcedo (2016).

Definition 2.1 *A distribution of play* is a function $h : Y \times X \times E \to [0,1]$ such that $\Pr(\mathbf{y} = y|\mathbf{x} = x, \mathbf{e} = e) = h(y, x, e)$. Let $\mathcal{H}$ be the set of all possible h.

The distribution of play contains information about the behavior of players. For example, whether they play according to NE in pure (PNE) or mixed strategies, how they mix among different equilibria in the region of multiplicity. Moreover, the distribution of play can be informative about the structure of the game, e.g., whether entry is simultaneous or sequential. When the researcher has particular behavioral assumptions in mind, say NE, then the distribution of play coincides with the predictions of the model when the equilibrium is unique and is closely related to the notion of selection mechanism in regions of multiplicity; see Kashaev and Salcedo (2016) for a more detailed discussion. The distribution of play is unknown to the econometrician.

Example 2.1 Suppose that the above entry game is a simultaneous move game of complete information; $\tilde{x}_i\tilde{\beta}_i = \tilde{\beta}_i > 0$, $i = 1, 2$. Assume, moreover, that firms always play a pure strategy NE and always play $y = (1, 0)$ if it is possible. Then,

$$h((0,0), x, e) = \begin{cases} 1 & \text{if } e_i > \bar{\beta}_i^{\intercal}\bar{x}_i, \quad i = 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

$$h((1,0), x, e) = \begin{cases} 1 & \text{if } e_1 \leq \bar{\beta}_i^{\intercal}\bar{x}_1 \text{ and } e_2 \geq \bar{\beta}_2^{\intercal}\bar{x}_2 - \tilde{\beta}_2 \\ 0 & \text{otherwise} \end{cases}$$

$$h((1,1), x, e) = \begin{cases} 1 & \text{if } e_i < \bar{\beta}_i^{\intercal}x_i - \tilde{\beta}_i, \quad i = 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

The definition of the distribution of play does not require the use of any particular equilibrium/solution concept. For instance, $h(y, x, e) = 1/4$ for all y, x and e is a valid distribution of play. However, it has nothing to do with the NE assumption.

2.2. SOR and NE constraints

Let $\theta = (\beta, h) \in \Theta = B \times \mathcal{H}$ be the parameter of the model. Then the conditional distribution of outcomes implied by the model is

$$p(y|x;\theta) = \int_E h(y,x,e) f_{e|x}(e|x,\beta) de$$

The set of possible distributions of play is too big. Example 2.1 shows that h does not have to be differentiable or even continuous. There is no connection between payoffs and outcomes. In this section I (i) show how relatively weak assumption on behavior of players, namely SOR, can be used to construct a conditional density implied by a smooth h, and (ii) construct a system of constraints on parameters that are equivalent to the NE assumption.

2.2.1. SOR constraints

Assume that the firms are maximizing expected payoffs and mutual knowledge of such behavior. I do not specify whether firms make their choices simultaneously or sequentially.

For given x and for any $\beta \in B$, let $A(y,x,\beta) \subseteq E$ be the set of all possible e such that y is the only outcome that survives two rounds of strictly dominant strategies elimination. Let $A^M(x,\beta) = \mathbb{R}^2 \setminus \bigcup_{y \in Y} A(y,x,\beta)$ be the region where the SOR assumption does not give unique predictions about the behavior of players.[8]

Then, the SOR assumption can be defined in terms of primitives of the model as follows.

Definition 2.2 A point $\theta \in \Theta$ is consistent with SOR if and only if

$$h(\mathbf{y},\mathbf{x},\mathbf{e}) \in \begin{cases} \{1\} & \text{if } \mathbf{e} \in A(\mathbf{y},\mathbf{x},\beta), \\ [0,1] & \text{if } \mathbf{e} \in A^M(\mathbf{x},\beta), \\ \{0\} & \text{otherwise} \end{cases} \qquad \text{a.s.} \tag{2.2}$$

[8]See the appendix for formal definition of these sets.

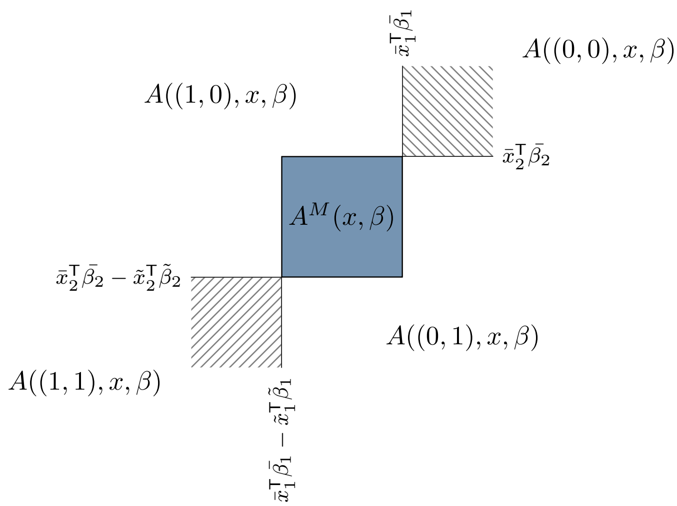


Figure 2.1 – Possible outcomes of the game under the SOR assumption.

In order for the domain of h not to depend on β, I will use

$$X \times E' = X \times \left\{ e \in E \mid \inf_{x \in X, \beta \in B} (\bar{x}_i^\intercal \bar{\beta}_i - \tilde{x}_i^\intercal \tilde{\beta}_i) \leq e_i \leq \sup_{x \in X, \beta \in B} \bar{x}_i^\intercal \bar{\beta}_i, i = 1, 2 \right\}$$

as the effective domain of h.[9]

Under the SOR assumption the conditional density implied by parameter θ is:

$$p(y|x;\theta) = \int_{A(y,x,\beta)} f_{e|x}(e|x;\beta)de + \int_{A^M(x,\beta)} h(y,x,e) f_{e|x}(e|x;\beta)de. \qquad (2.3)$$

2.2.2. NE constraints

Every NE is second order rational. So, the distribution of play implied by the NE assumption can be different from the distribution of play implied by the SOR assumption *only* when $e \in A^M(x,\beta)$. By Assumption 2 $A^M(x,\beta)$ has a positive measure with probability 1. The set $A^M(x,\beta)$ is a region of multiplicity of equilibria under the NE assumption. There are three NE: two in pure and one in mixed strategies. For fixed x, e and β such that $e \in A^M(x,\beta)$, let $\alpha(y,x,e,\beta)$ be the probability that y is played in the mixed strategy NE.[10] Denote $\alpha(x,e,\beta) = (\alpha(y,x,e,\beta))_{y \in Y}$ and $h(x,e) = (h(y,x,e))_{y \in Y}$. Since I am not assuming the way players randomize among

[9] Assumption that the domain of h is $X \times E'$ is not restrictive, since $A^M(x,\beta)$ is a strict subset of E' for all x and β.

[10] See the appendix for a formal definition of α.

different NE, $h(x, e)$ can be equal to any convex combination of $\alpha(x, e, \beta)$ (mixed NE), $(0,0,1,0)^\intercal$ ("entry-nonentry" NE) and $(0,0,0,1)^\intercal$ ("nonentry-entry" NE).

Definition 2.3 A pair $\theta_{\text{NE}} = (\beta_{\text{NE}}, h_{\text{NE}}) \in \Theta$ is consistent with the NE assumption if for almost all $(x^\intercal, e^\intercal)^\intercal \in X \times E$, θ_{NE} is consistent with SOR, and if $e \in A^M(x, \beta_{\text{NE}})$, then

$$h_{\text{NE}}(x, e) \in \text{co}\Big((0,0,1,0)^\intercal, (0,0,0,1)^\intercal, \alpha(x, e, \beta_{\text{NE}})\Big), \tag{2.4}$$

where $\text{co}(A)$ denotes a convex hull of the set A.

Let $\text{NE} \subseteq \Theta$ be the set of all θ that are consistent with NE.

Proposition 2.1 *Under Assumptions 1 and 2, there are* known *functions* $m_j : \Theta \to \mathbb{R}$, $j = 0, 1, 2, 3$*, such that for every* $j = 0, 1, 2, 3$ *(i)* m_j *continuously differentiable in* β*; (ii)* m_j *is affine in* h*; (iii)*

$$\theta \in \text{NE} \iff \begin{cases} m_0(\theta) = 0, \\ m_1(\theta) \geq 0, \\ m_2(\theta) \geq 0, \\ m_3(\theta) \geq 0. \end{cases}$$

Proof. See the appendix ■

The idea behind Proposition 2.1 is simple. The convex hull of NE can be characterized as an intersection of one hyperplane, $m_0(\theta) = 0$, and three half spaces, $m_j(\theta) = 0$, $j = 1, 2, 3$. If none of the inequality constraints is binding then the firms play every NE with positive probability; if m_1 is binding then agents never play "entry-nonentry" NE; if m_2 is binding then agents never play "nonentry-entry" NE; if m_3 is binding then agents never play mixed strategy NE. Note that at most two inequality constraints can be binding simultaneously.

2.3. Parameter space and identified set

In order to have nice approximating results I am assuming a certain degree of smoothness of the distribution of play in the region of multiplicity.

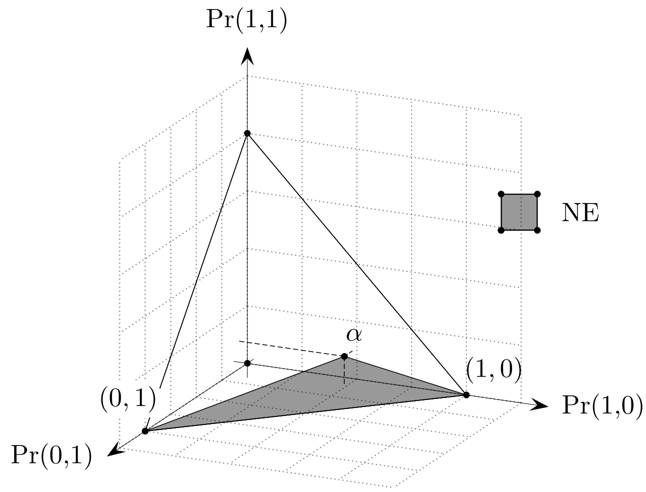


Figure 2.2 – Predictions for NE in the multiplicity region. The point α corresponds to the mixed NE. The gray region is characterized by an intersection of three half spaces and one hyperplane.

Let $g : M \times Z \to \mathbb{R}$, where $Z \subseteq \mathbb{R}^{d_z}$ and M is a finite set. For λ a d_z-dimensional vector of nonnegative integers, let $|\lambda| = \sum_{i=1}^{d_z} \lambda_i$, $D^\lambda g(m,z) = \partial^{|\lambda|} g(m,z)/\partial z_1^{\lambda_1} \dots \partial z_{d_z}^{\lambda_{d_z}}$. Then, for κ and $\kappa_0 > 0$, I define the following modification of norms, which are used in Santos (2012):

$$\|g\|_s = \sqrt{\max_{m\in M} \sum_{|\lambda|\le \kappa+\kappa_0} \int_Z [D^\lambda g(m,z)]^2 dz},$$
$$\|g\|_c = \max_{m\in M} \max_{|\lambda|\le\kappa} \sup_{z\in Z} \left| D^\lambda g(m,z) \right|.$$

Denote the corresponding spaces by

$$W^s(M \times Z) = \{g : M \times Z \to \mathbb{R} \mid \|g\|_s < \infty\},$$
$$W^c(M \times Z) = \{g : M \times Z \to \mathbb{R} \mid \|g\|_c < \infty\}.$$

Assumption 3 The parameter space is $\Theta = B \times \mathcal{H}$, where B is a compact subset of $\mathbb{R}^{d_\beta}$ and

$$\mathcal{H} = \left\{ h \in W^s(Y \times X \times E') \mid \|h\|_s \le K, 0 \le h \le 1, \sum_y h(y,\cdot,\cdot) = 1 \right\},$$

where $\min\{\kappa, \kappa_0\} > d_x/2 + 1$, and K is a known finite constant.

Note that Θ is compact under $\|\theta\|_c = \|\beta\|_e + \|h\|_c$ (see Lemma A.2. in Santos

(2012)).

Definition 2.4 The identified set is

$$\Theta_0 = \{\theta \in \Theta \mid p(\mathbf{y}|\mathbf{x};\theta) = p_0(\mathbf{y}|\mathbf{x}) \text{ a.s.}\}\,, \tag{2.5}$$

where $p_0(y|x) = p(y|x;\theta_0)$ is the true conditional distribution.

Usually the identified set is not a singleton.[11] However, conditional densities implied by different points in the identified set coincide. Define the squared Pearson pseudo distance on the space of probability densities with respect to a dominating σ-finite positive measure μ:

$$\chi^2(P_1, P_2) = \int \left(\frac{P_1}{P_2} - 1\right)^2 P_2 d\mu$$

Note that the true joint probability density of $(\mathbf{y}, \mathbf{x})$ is $P_0(y, x) = p_0(y|x)f_x(x)$. Sometimes I will abuse notation and use $\chi(\theta_1, \theta_2)$ instead of $\chi(p(\theta_1), p(\theta_2))$.

An equivalent definition of the identified set is:

$$\Theta_0 = \{\theta \in \Theta \mid \chi(\theta, \theta_0) = 0\}\,,$$

Let $\lambda_{\min}(\theta_0)$ be the minimal eigenvalue of the matrix

$$\Lambda(\theta_0) = \mathbb{E}\Big\{\partial_\beta \log(p(\mathbf{y}|\mathbf{x}, \beta, h_0))|_{\beta=\beta_0} \;\; \partial_{\beta^\intercal} \log(p(\mathbf{y}|\mathbf{x}, \beta, h_0))|_{\beta=\beta_0}\Big\}$$

Assumption 4 (Rank Condition)

$$\inf_{\theta_0 \in \Theta_0 \bigcap \mathrm{NE}} \lambda_{\min}(\theta_0) > 0. \tag{2.6}$$

Assumption 4 is a rank condition and implies local point identification of β_0 when h_0 is known. It is important that the derivative of the log-likelihood is taken with h_0 being fixed. Existence of player specific covariates with large but bounded support is sufficient for Assumption 4. (See the appendix for a formal statement.)

[11] Point identification of both β and h can be achieved if one assumes existence of player specific excluded covariates with unbounded support that enter h only through preferences; see Kashaev and Salcedo (2016).

Under Assumption 4, $\Lambda(\theta_0)$ induces a norm

$$\|\beta\|_{\theta_0} = \sqrt{\beta^{\intercal}\Lambda(\theta_0)\beta}.$$

in $\mathbb{R}^{d_\beta}$ for every $\theta_0 \in \Theta_0 \bigcap \text{NE}$.

2.4. Sieve maximum likelihood setting

The sieve space for h takes the form

$$\mathcal{H}_{k(n)-d_\beta} = \left\{h \in \mathcal{H} \mid h(y,e,x) = \Pi_y^{\intercal}\psi^{J(n)}(e,x)\right\},$$

where Π is a vector of unknown sieve coefficients, $\psi^{J(n)}$ is a vector of known basis functions, such as polynomial series or splines, that are at least $[d_x/2]+1$ times continuously differentiable and $k(n) = 4J(n)^d + d_\beta$ is the dimensionality of $\Theta_{k(n)} = B \times \mathcal{H}_{k(n)-d_\beta}$.

Let γ_n be the sieve approximation rate. That is, for every $h \in \mathcal{H}$ there exists $h_k \in \mathcal{H}_{k(n)-d_\beta}$ such that for some c_1

$$\|h_k - h\|_\infty \leq c_1\gamma_n. \tag{2.7}$$

Assumption 5 (Sieve Spaces) (i) $\{\beta \in B : (\beta, h) \in \Theta_0\} \subseteq int(B)$, where $int(B)$ is an interior of B.

(ii) (i) for each $k \geq 1$, Θ_k is closed under $\|\cdot\|_c$ with $dim(\Theta_k) < \infty$; (ii) $\emptyset \neq \Theta_k \subseteq \Theta_{k+1} \subseteq \Theta$ for all $k \geq 1$, and $\overline{\bigcup_k \mathcal{H}_k}$ is dense in $\mathcal{H}$ under $\|\cdot\|_c$.

(iii) $\log N(\epsilon, \Theta_n, \|\cdot\|_c) = o(n)$ for every $\epsilon > 0$, where $N(\epsilon, \Theta_n, \|\cdot\|_c)$ is a covering number, and $k(n) = o(n)$.

Assumption 5.(i) ensures that I can apply Taylor's theorem to p and the NE constraints with respect to β around points in the identified set. Assumptions 5.(ii) and 5.(iii) are standard assumptions in the sieve literature and are satisfied by many sieves.

Assumption 6 (Data) The data $\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1}^n$ is an i.i.d. random sample from a unique density $p_0 f_x$;

The following theorem presents consistency in the one sided Hausdorff metric, rates of convergence in terms of the Pearson distance and $\|\cdot\|_{\theta_0}$. Let L_n be a conditional sample log-likelihood, that is, $L_n(\theta) = \sum_{i=1}^n \log p(\mathbf{y}_i|\mathbf{x}_i;\theta)$.

Theorem 2.2 *Let $\hat{\Theta}_n \subseteq \Theta_{k(n)}$ and $\tilde{\Theta}_n \subseteq \Theta_{k(n)}$ be collections of $\hat{\theta}_n = (\hat{\beta}_n, \hat{h}_n)$ and $\tilde{\theta}_n = (\tilde{\beta}_n, \tilde{h}_n)$ respectively that satisfy*

$$L_n(\hat{\theta}_n) = \sup_{\theta\in\Theta_{k(n)}} L_n(\theta),$$

$$L_n(\tilde{\theta}_n) = \sup_{\theta\in\Theta_{k(n)}\bigcap \mathrm{NE}} L_n(\theta),$$

and $\delta_n = \max\left\{\left(\frac{k(n)}{n}\right)^{1/2}, \gamma_n\right\}$. Under Assumptions 1, 2, 3, 5 and 6:

(i) $\sup_{\hat{\theta}_n\in\hat{\Theta}_n} \inf_{\theta_0\in\Theta_0} \left\|\hat{\theta}_n - \theta_0\right\|_c = o_p(1)$;

(ii) $\chi(\hat{\theta}_n, \theta_0) = O_p(\delta_n)$ *for all* $\hat{\theta}_n \in \hat{\Theta}_n$ *and* $\theta_0 \in \Theta_0$.

If, moreover, $\Theta_0 \bigcap \mathrm{NE} \neq \emptyset$, Assumption 4 is satisfied, and $\delta_n \log\log(n) = o_p(n^{-1/4})$, then

(iii)

$$\sup_{\tilde{\theta}_n\in\tilde{\Theta}_n} \inf_{\theta_0\in\Theta_0\bigcap \mathrm{NE}} \left\|\tilde{\beta}_n - \beta_0\right\|_{\theta_0} = o_p(n^{-1/4}). \tag{2.8}$$

Proof. See Appendix. ∎

Results 1 and 2 are applications of Theorem 3.1 of CTT. The last result is essential for testing the Nash constraints. It allows to linearize the constraints in β which together with affinity in h provides the asymptotic distribution of the test statistics introduced in the next section.

2.5. Sieve LR type statistics

Recall that $\mathrm{NE} = \{\theta \in \Theta : m_0(\theta) = 0, m_1(\theta) \geq 0, m_2(\theta) \geq 0, m_3(\theta) \geq 0\}$. I want to test the null hypothesis that $\Theta_0 \bigcap \mathrm{NE} \neq \emptyset$ against the alternative that $\Theta_0 \bigcap \mathrm{NE} = \emptyset$.

Define the sieve LR statistic as follows.

$$T_{n,0} = 2\left[\sup_{\theta\in\Theta_{k(n)}} L_n(\theta) - \sup_{\theta\in\Theta_{k(n)}\cap \mathrm{NE}} L_n(\theta)\right].$$

Since NE involves inequality constraints, even in the parametric point identified case, the limiting null distribution depends on which inequality constraints are binding at θ_0.

Define

$$\begin{aligned}
\mathrm{NE}_j &= \mathrm{NE}\bigcap\Big\{\theta\in\Theta : m_j(\theta) = 0\Big\}, \quad j = 1,2,3,\\
\mathrm{NE}_{i,j} &= \mathrm{NE}\bigcap \mathrm{NE}_i \bigcap \mathrm{NE}_j, \quad (i,j)\in\{(1,2),(1,3),(2,3)\}.
\end{aligned}$$

The sets NE_1, NE_2 and NE_3 contain $\theta\in\Theta$ such that "entry-nonentry", "nonentry-entry" and the mixed Nash equilibrium are never played in the multiplicity region respectively. $\mathrm{NE}_{1,2}$, $\mathrm{NE}_{1,3}$ and $\mathrm{NE}_{2,3}$ are the sets of $\theta\in\Theta$ such that the mixed, "nonentry-entry" and "entry-nonentry" Nash equilibrium is always played in the multiplicity region.

Define the following sieve LR type statistics:

$$\begin{aligned}
T_{n,j} &= 2\left[\sup_{\theta\in\Theta_{k(n)}} L_n(\theta) - \sup_{\theta\in\Theta_{k(n)}\cap \mathrm{NE}_j} L_n(\theta)\right], j = 1,2,3,\\
T_{n,i,j} &= 2\left[\sup_{\theta\in\Theta_{k(n)}} L_n(\theta) - \sup_{\theta\in\Theta_{k(n)}\cap \mathrm{NE}_{i,j}} L_n(\theta)\right], (i,j)\in\{(1,2),(1,3),(2,3)\}.
\end{aligned}$$

As one can see, each of the statistics corresponds to the case when one or two NE inequality constraints are binding. The following theorem states that under different nulls, the above statistics have a tight limit.

Theorem 2.3 *Under the assumptions of Theorem 2.2,*

(i) If $\Theta_0\bigcap \mathrm{NE}_{i,j}\neq\emptyset$ for some $(i,j)\in\{(1,2),(1,3),(2,3)\}$, then $T_{n,i,j}\to_d \chi^2(3)$;

(ii) If $\Theta_0\bigcap \mathrm{NE}_{i,j}=\emptyset$ for all $(i,j)\in\{(1,2),(1,3),(2,3)\}$ and $\Theta_0\bigcap \mathrm{NE}_j\neq\emptyset$ for some $j\in\{1,2,3\}$, then $T_{n,j}$ has a tight limit;

(iii) If $\Theta_0\bigcap \mathrm{NE}_j=\emptyset$ for all $j=1,2,3$ and $\Theta_0\bigcap \mathrm{NE}\neq\emptyset$, then $T_{n,0}$ has a tight limit.

Note that in Theorem 2.3, I partitioned the original null hypothesis into the finite set of mutually exclusive hypotheses:

$$\left[\Theta_0 \bigcap \mathrm{NE} \neq \emptyset\right] \Leftrightarrow \begin{cases} \left[\Theta_0 \bigcap \mathrm{NE}_{1,2} \neq \emptyset\right], \\ \left[\Theta_0 \bigcap \mathrm{NE}_{1,3} \neq \emptyset\right], \\ \left[\Theta_0 \bigcap \mathrm{NE}_{2,3} \neq \emptyset\right], \\ \left[\Theta_0 \bigcap \mathrm{NE}_1 \neq \emptyset\right] \ \& \ \left[\forall\,(i,j) \in \{(1,2),(1,3),(2,3)\}, \Theta_0 \bigcap \mathrm{NE}_{i,j} = \emptyset\right], \\ \left[\Theta_0 \bigcap \mathrm{NE}_2 \neq \emptyset\right] \ \& \ \left[\forall\,(i,j) \in \{(1,2),(1,3),(2,3)\}, \Theta_0 \bigcap \mathrm{NE}_{i,j} = \emptyset\right], \\ \left[\Theta_0 \bigcap \mathrm{NE}_3 \neq \emptyset\right] \ \& \ \left[\forall\,(i,j) \in \{(1,2),(1,3),(2,3)\}, \Theta_0 \bigcap \mathrm{NE}_{i,j} = \emptyset\right], \\ \left[\Theta_0 \bigcap \mathrm{NE} \neq \emptyset\right] \ \& \ \left[\forall\, j \in \{1,2,3\}, \Theta_0 \bigcap \mathrm{NE}_j = \emptyset\right]. \end{cases}$$

When two of the inequality constraints are binding, say $m_2(\theta_0) = 0$ and $m_3(\theta_0) = 0$, then agents always play "entry-nonentry" NE in the multiplicity region. Hence, in the multiplicity region $h_0((1,0), x, e) = 1$, the model becomes point identified and there are 3 binding constraints. The asymptotic null distribution of $T_{n,j}$ is complicated. In the point identified case, similar to $T_{n,i,j}$, it reduces to a chi-squred distribution with 2 ($j = 1,2,3$) and 1 ($j = 0$) degrees of freedom. In the next section I present a computationally simple approach to compute critical values of the corresponding limiting distributions based on the multiplier bootstrap. Once the bootstrap validity is established, one can test for NE by the following procedure: let $c(1-\alpha)$ and $c_{n,j}(1-\alpha)$ be the $(1-\alpha)$ quantiles of $\chi^2(3)$ and the multiplier bootstrap distribution.

Step 1. Compute $T_{n,1,2}$, $T_{n,1,3}$, and $T_{n,2,3}$. If $\min\{T_{n,1,2}, T_{n,1,3}, T_{n,2,3}\} \leq c(1-\alpha)$, then accept the null hypothesis. Otherwise proceed to Step 2.

Step 2. Compute $T_{n,i}$, $i = 1,2,3$. If $T_{n,i} \leq c_{n,i}(1-\alpha)$ for some $i = 1,2,3$, then accept the null hypothesis. Otherwise proceed to Step 3.

Step 3. Compute $T_{n,0}$. If $T_{n,i} \leq c_{n,0}(1-\alpha)$, then accept the null hypothesis. Otherwise the null is rejected.

The procedure is conservative. However, if NE is rejected at Step 1, then the players are likely to always mix between different NE; if NE is rejected at Step 2, then the players are likely to always mix between all NE.[12] Thus, as an intermediate step one can learn what agents do in the region of multiplicity.

[12]The order of steps is not important since the null hypotheses in Theorem 2.3 are mutually exclusive.

2.6. Multiplier bootstrap

Once the asymptotic null distribution of each statistic is derived, I show consistency of the multiplier bootstrap with i.i.d weights. Consider, for instance, the asymptotic null distribution of $T_{n,0}$. Informally the proposed procedure is as follows. (i) Generate R_n samples of size n of positive weights from a standard exponential distribution with mean and variance equal to 1; (ii) For each bootstrap sample compute a “weighted” version of the statistic, $T_{n,0}^{w,r}$; (iii) Use

$$c_{n,0}(1-\alpha) = \inf\left\{\tau : \frac{1}{R_n}\sum_{r=1}^{R_n} \mathbb{1}(T_{n,0}^{w,r} \leq \tau) \geq 1-\alpha\right\}, \tag{2.9}$$

as a critical value in the corresponding step of the procedure described in the previous section.

The bootstrap weights satisfy the following assumption.

Assumption 7 $\{\mathbf{w}_i\}$ is a positive, i.i.d. sequence drawn from the distribution of a positive random variable $\mathbf{w}$ with $\mathbb{E}[\,\mathbf{w}\,] = 1$, $\mathbb{E}[\,(\mathbf{w}-1)^2\,] = 1$, $\int_0^{+\infty}\sqrt{\Pr(|\mathbf{w}-1| \geq t)}dt < \infty$ and independent of $\{(\mathbf{y}_i, \mathbf{x}_i)\}_{i=1}^n$.

Let $\hat{\theta}_n$ be an unconstrained maximizer of $L_n(\theta)$. Define

$$\begin{aligned}
\mathrm{NE}(\hat{\theta}_n) &= \{\theta \in \Theta : m_0(\theta) = m_0(\hat{\theta}_n)\},\\
\mathrm{NE}_j(\hat{\theta}_n) &= \mathrm{NE}(\hat{\theta}_n)\bigcap\left\{\theta \in \Theta : m_j(\theta) = m_j(\hat{\theta}_n)\right\}, \quad j = 1,2,3,\\
\mathrm{NE}_{i,j}(\hat{\theta}_n) &= \mathrm{NE}(\hat{\theta}_n)\bigcap \mathrm{NE}_i(\hat{\theta}_n)\bigcap \mathrm{NE}_j(\hat{\theta}_n), \quad (i,j) \in \{(1,2),(1,3),(2,3)\}.
\end{aligned}$$

and

$$\begin{aligned}
L_n^w(\theta) &= \sum_{i=1}^n \mathbf{w}_i \log\left(p(\mathbf{y}_i|\mathbf{x}_i,\theta)\right),\\
T_{n,0}^w &= 2\left[\sup_{\theta\in\Theta_{k(n)}} L_n^w(\theta) - \sup_{\theta\in\Theta_{k(n)}\cap\mathrm{NE}(\hat{\theta}_n)} L_n^w(\theta)\right],\\
T_{n,j}^w &= 2\left[\sup_{\theta\in\Theta_{k(n)}} L_n^w(\theta) - \sup_{\theta\in\Theta_{k(n)}\cap\mathrm{NE}_i(\hat{\theta}_n)} L_n^w(\theta)\right], j = 1,2,3,\\
T_{n,i,j}^w &= 2\left[\sup_{\theta\in\Theta_{k(n)}} L_n^w(\theta) - \sup_{\theta\in\Theta_{k(n)}\cap\mathrm{NE}_{i,j}(\hat{\theta}_n)} L_n^w(\theta)\right], (i,j) \in \{(1,2),(1,3),(2,3)\}.
\end{aligned}$$

In contrast to CTT and Tao (2014), the bootstrap statistics are not properly centered since one cannot guarantee that the NE constraints will be satisfied at $\hat{\theta}_n$. As a result, the bootstrap critical values can be conservative. For bootstrap validity I need to strengthen the rank condition. Recall that $\lambda_{\min}(\theta_0)$ is the minimal eigenvalue of the matrix

$$\Lambda(\theta_0) = \mathbb{E}\Big\{\partial_\beta \log(p(\mathbf{y}|\mathbf{x},\beta,h_0))|_{\beta=\beta_0} \;\; \partial_{\beta^\intercal} \log(p(\mathbf{y}|\mathbf{x},\beta,h_0))|_{\beta=\beta_0}\Big\}$$

Assumption 8 (Rank Condition 2)

$$\inf_{\theta_0\in\Theta_0} \lambda_{\min}(\theta_0) > 0. \tag{2.10}$$

Theorem 2.4 *Under the assumptions of Theorem 2.3, Assumption 7 and Assumption 8,*

(i) if $\Theta_0 \bigcap \mathrm{NE}_{i,j} = \emptyset$ *for all* $(i,j) \in \{(1,2),(1,3),(2,3)\}$ *and* $\Theta_0 \bigcap \mathrm{NE}_j \neq \emptyset$ *for some* $j \in \{1,2,3\}$*, then for* $c_{n,j}(1-\alpha)$ *defined in (2.9)*

$$\lim \Pr\left(T_{n,j} \leq c_{n,j}(1-\alpha)\right) \geq 1-\alpha.$$

(ii) if $\Theta_0 \bigcap \mathrm{NE}_j = \emptyset$ *for all* $j = 1,2,3$ *and* $\Theta_0 \bigcap \mathrm{NE} \neq \emptyset$*, then*

$$\lim \Pr\left(T_{n,0} \leq c_{n,0}(1-\alpha)\right) \geq 1-\alpha.$$

(iii) if $\Theta_0 \bigcap \mathrm{NE} = \emptyset$*, then the proposed testing procedure rejects the NE assumption with probability approaching 1.*

2.7. Monte Carlo experiment and empirical applcation

In this section I provide results for Monte Carlo experiments and empirical application based on the model and the data used in Grieco (2014).

2.7.1. Monte Carlo experiment

I provide results for the finite sample performance of my testing procedure (sLR). I also test the NE assumption by using BMM approach coupled with Andrews and Shi (2013)(BMM+AS). I consider a two player entry game of complete information without covariates.[13] I assume that the conditional distribution of payoffs is known to the econometrician. However, the distribution of play is not assumed to be known.

I assume that the econometrician knows the distribution of payoffs for the following reason. If the payoff parameters are unknown, then generically β_0 is not pointidentified. As a result, it is not clear how to construct a data generating process (DGP) that cannot be rationalized by the NE assumption. Thus, the power properties of the testing procedure may not be reliable. This is not the case when β_0 is known.[14]

The payoff of player i for an outcome $y \in Y$ is given by:

$$u_i(y, e) = (-y_{-i} - e_i)y_i,$$

where $\mathbf{e} \sim N(0, I)$.

For the above payoff specification the multiplicity region is $[-1, 0]^2$. I consider 3 different DGPs under the null that the NE assumption is correct. In the multiplicity region agents (i) always play the mixed strategy NE (Mixed); (ii) play "entry-nonentry" , "nonentry-entry" and the mixed strategy NE with probabilities 0.3, 0.3 and 0.4 respectively (All); (iii) given (e_1, e_2), play "entry-nonentry" and "nonentry-entry" NE with probabilities $\phi(e_1, e_2)$ and $(1 - \phi(e_1, e_2))$ respectively, where $\phi : [-1, 0]^2 \to [0, 1]$ such that

$$\phi(e_1, e_2) = \frac{e^{(e_2 - e_1)} - e^{-1}}{e - e^{-1}}.$$

The last DGP has the following justification. Since the expected payoff from playing the mixed strategy NE is dominated by the monopoly payoff, one can assume that firms play only pure strategy NE (PNE). Assume that the players bargain about which PNE has to be played. The bargaining power of player 1 is captured by $\phi(\cdot, \cdot)$. Then for a given realization of payoffs, players choose the probability that agent 1 is

[13] The case with covariates is in progress.

[14] I conducted MC simulations assuming that β_0 is unknown under the null that the NE assumption is satisfied. The results are very similar to the case with known β_0.

the monopolist, p^*, according to the following maximization problem.

$$\max_{p\in[0,1]} \Big[pu_1((1,0),e)\Big]^{\phi(e)} \Big[(1-p)u_2((0,1),e)\Big]^{1-\phi(e)}$$

So, $p^* = \phi(e)$.

Under the alternative that the NE assumption is false I consider 5 different DGPs. Let $\alpha(y,e)$ be the probability that y is played under the mixed strategy NE in the multiplicity region. Assume that with probability $(1-p_{NN})$ agents play the mixed strategy NE and with probability p_{NN} they never enter in the multiplicity region. That is, for $e \in [-1,0]^2$,

$$\Pr(\mathbf{y} = y|\mathbf{e} = e) = (1-p_{NN})\alpha(y,e) + p_{NN}\mathbb{1}(y = (0,0)),$$

where $p_{NN} \in \{0.2, 0.4, 0.6, 0.8, 1\}$.

I use second order polynomials to approximate the unknown distribution of play.[15] To impose the NE constraints, I generate 100 Halton points on the multiplicity region, $[-1,0]^2$, and evaluate every NE constraint at every Halton point.[16] The number of bootstrap replications is equal to 500. I end up with 18 parameters (6 parameters for each $h(y)$, $y \in \{(0,0),(1,1),(1,0)\}$) and from 0 to 200 equality constraints and from 200 to 700 inequality constraints depending on which statistic I am computing.

The experiment is run for two sample sizes: $n = 500$ and $n = 1000$. For each sample size, 1000 such samples are generated. The result of the simulations for the first three DGPs are displayed in tables 1 and 2.

Table 2.1 – Percent of rejections in MC experiment, $n = 500$

	Mixed		All		Bargaining	
α,%	sLR	BMM+AS	sLR	BMM+AS	sLR	BMM+AS
20	12.3	1.1	10	0.1	5.1	2.5
10	3.8	0.5	4.2	0	1.8	0.4
5	1.8	0.1	2	0	0.5	0

Notes: α is a significance level; $\#MC = 1000$, $\#\mathbf{bootstrap} = 500$.

[15] The results for third order polynomials are qualitatively the same.

[16] For details on Halton sequences, see Bhat (2001). For some DGPs I used 200 and 500 Halton points. The results are approximately the same.

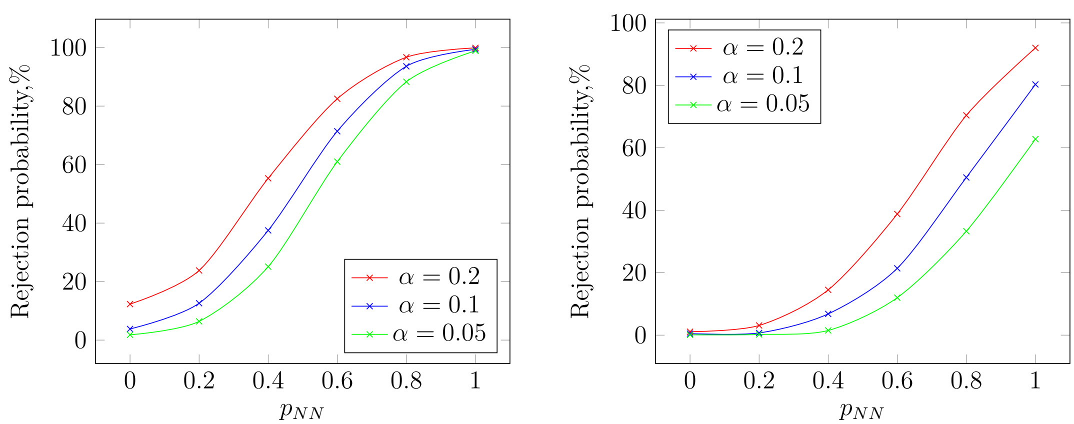
Table 2.2 – Percent of rejections in MC experiment, $n = 1000$

	Mixed		All		Bargaining	
α,%	sLR	BMM+AS	sLR	BMM+AS	sLR	BMM+AS
20	14	1.1	14.7	0	7.6	3.8
10	6.1	0.4	7.5	0	3.4	0.7
5	3	0.3	3.5	0	1.1	0.3

Notes: α is a significance level; $\#MC = 1000$, $\#\textbf{bootstrap} = 500$.

Both approaches are conservative. However, the proposed sieve LR procedure is substantially less conservative. It is worth to note that the procedure based on BMM with AS is not designed to test the NE assumption, it is uniform, and does not assume smoothness of the distribution of play. I believe that with a proper adjustment of tuning parameters the performance of the "BMM+AS" procedure can be improved.

The power results are presented in figures 3 and 4. The power of both procedures improves with an increase in sample size. The "sLR" procedure has better power than the "BMM+AS" procedure.

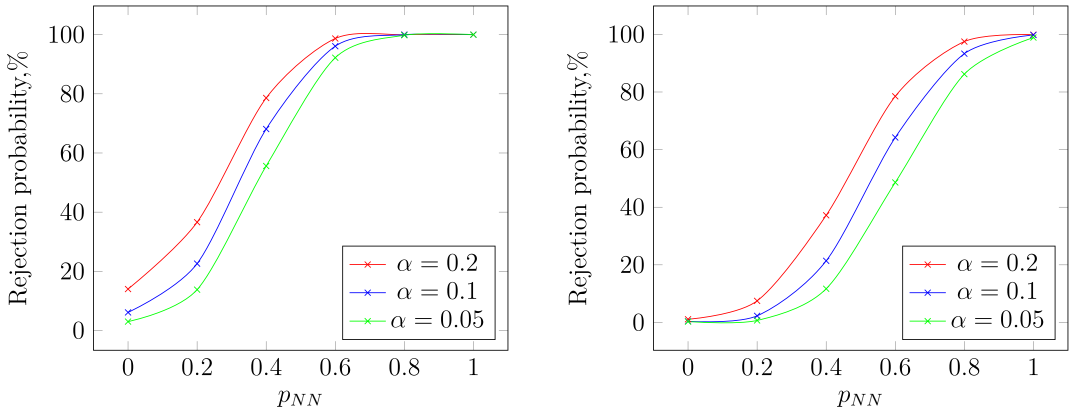


(a) sLR **(b)** BMM+AS

Figure 2.3 – Power curves, $n = 500$

2.7.2. Empirical application

In Grieco (2014) author presents an important entry model allowing for both private and public information. He applied his approach to study entry and exit



(a) sLR **(b)** BMM+AS

Figure 2.4 – Power curves, $n = 1000$

patterns of small rural grocery stores in the US between 1998 and 2002. I consider a complete information version of that model assuming that the unobserved payoff shifters are perfectly correlated.[17] Perfectly correlated unobserved payoff shifters imply that there is only market specific heterogeneity.

The data I use contains for each market information on whether the firm was present in 2002, the population, the distance from a supercenter (Walmart), and whether the firm was active in 1998. In total there are 4803 observations.[18] Profits of firm i can be written as,

$$u_i(y, x, e, \beta) = (\mu(x, \beta) - \delta(x, \beta)y_{-i} - e)y_i,$$

where

$$\begin{aligned}\mu(x, \beta) =& \beta_0 + \beta_1 \mathbb{1}(pop > 3k) + \beta_2 \mathbb{1}(pop > 6k) + \\ & \beta_3 \mathbb{1}(supercenter < 20mi) + \beta_4 \mathbb{1}(iInactive1998),\end{aligned}$$

$$\delta(x, \beta) = \beta_5 + \beta_6 \mathbb{1}(pop > 3k) + \beta_7 \mathbb{1}(pop > 6k) + \beta_8 \mathbb{1}(supercenter < 20mi).$$

and $\mathbf{e}|(\mathbf{x} = x) \sim N(0, 1)$.

I use fifth order degree polynomials to approximate the distribution of play. To

[17] I estimated the model assuming that correlation between unobserved payoff shifters is unknown. The value of correlation parameter that maximizes both the SOR and the NE sample log-likelihood objective functions becomes equal to 1.

[18] This data was graciously provided by Paul Grieco. For more details on the data set see Grieco (2014).

impose the NE constraints, I generate 60 Halton points for every value of covariates. The values of the sieve ML objective functions are presented in table 3.

Table 2.3 – Sample log-likelihood

	SOR	All	no (1,0)	no (0,1)	no mixed	mixed only
Log-likelihood	-2974	-3099	-3443	-3380	-10250	-3580

Notes: "SOR" - the value of log-likelihood under the SOR assumption; "All" - all NE can be played; "no (1,0)" - "entry-nonentry" equilibrium is never played; "no (0,1)" - "nonentry-entry" equilibrium is never played; "no mixed" - the mixed strategy equilibrium is never played; "mixed only" - the mixed strategy equilibrium is always played; Sample size= 4803,#**bootstrap** = 200.

I use 200 bootstrap replications to approximate the critical values. The NE assumption is rejected at the 5%-significance level.

2.8. Summary

This paper considers the problem of testing the *Nash equilibrium* assumption in partially identified semiparametric entry games of complete information. The procedure allows for nonparametric selection of NE in the regions of multiplicity. In principle, one can employ the proposed methodology to test for nested equilibrium assumptions as long as these assumptions can be represented by the set of equality/inequality constraints on parameters.

It is worth to note that the asymptotic results presented in the paper hold under a fixed underlying data generating process P_0. Since the asymptotic distribution depends on whether or not the NE inequality constraints are binding, the procedure may suffer from uniformity issues. I leave this important problem for the future work.

Chapter 3

Binary outcome models with pure-choice based data

Abstract In this paper I propose a generalized method of moments (GMM) type procedure to estimate parametric binary choice models when the researcher only observes the data with a particular response and has some information about the distribution of the covariates. An example might be where the firm observes characteristics of its own customers only and knows something about the distribution of the whole population of customers. This auxiliary information comes in the form of moments. I present an application based on the data on police-reported car accidents in Seattle. Publicly available information on the distribution of drivers' characteristics in Seattle allows me to estimate the probability of a two-car collision.

This paper considers identification and estimation of a parametric binary choice model with a pure choice-based data sample. To be more specific, let $(\mathbf{y}, \mathbf{x})$ satisfy $\mathbf{y} = \mathbb{1}(\mathbf{x}^T\theta_0 + \epsilon \geq 0)$, where $\mathbb{1}(\cdot)$ is an indicator function, ϵ is an error term assumed to be independent of $\mathbf{x}$, and θ_0 is the object of interest. The researcher observes the pair $(\mathbf{y}, \mathbf{x})$ only if $\mathbf{y} = 1$. I will call the sample drawn according to the above scheme a contaminated or pure choice-based sample. The simplest example of such a situation is when people choose whether to participate in a survey or not but the researcher only observes characteristics of those who decide to do so.

It is well known that the contaminated sample in itself does not identify the parameters of the conditional choice probabilities. To achieve identification, I need some independent source of information. Lancaster and Imbens (1996) proposed an efficient semiparametric procedure for estimating the binary choice model parameters when an additional random sample of covariates from the whole population is available. They impose the strong requirement that the additional sample contains information on *all* of the covariates. In practice, obtaining a sample from the whole population might be too costly or even impossible, some of the covariates may be observable only in the contaminated sample.

The above problem can be described in terms of the multinomial (MN) sampling scheme. In MN sampling, we partition the support of $\mathbf{y}$ into two disjoint sets (strata), namely $S_0 = \{0\}$ and $S_1 = \{1\}$. Then, with known probability p_i so that $p_0 + p_1 = 1$, we choose the strata S_i. Finally, we randomly draw an observation from the selected strata. I consider the extreme case of MN sampling when $p_1 = 1$ and $p_0 = 0$. Other interpretations I can think of are data sets with non-random attrition or models of missing data.[1] The key idea used in both cases is to weight the observed moment conditions or likelihood by the inverse probability of the later being observed. In the MN sampling literature these weights are defined by the nature of the sampling scheme. In data sets with non-random attrition or models of missing data, sometimes, to avoid the effects of attrition, panel data sets are augmented with new units randomly drawn from the original population, so-called refreshment samples. These refreshment samples are used to conduct the inverse probability weighting.

MN sampling has been widely studied in the literature. Manski and Lerman (1977) proposed a weighted maximum likelihood estimator in the context of choice-based sampling. Cosslett (1981a), Cosslett (1981b) and Imbens (1992) investigated the properties of the maximum likelihood and the Chamberlain (1987) type method of moments techniques, when stratification is based on exogenous variables. In a recent paper, Tripathi (2011) showed how to do efficient moment based inference using the generalized method of moments (GMM) under the assumption that aggregate shares are known.

In the context of panel data sets with non-random attrition, Hellerstein and Imbens (1999) analyzed the estimation of regression models with linear selection probability under moment restrictions that come from auxiliary data. Their moment restrictions allowed them to estimate weights that inflate the observed data so that the standard GMM estimator applied to the weighted data leads to consistent estimates. Nevo (2003) used Hellerstein and Imbens (1999)'s approach, but, instead of

[1] See Ridder (1992) and Hirano et al. (1998) for detailed discussions.

the linear selection probability model, he assumed the Logit one. Both Hellerstein and Imbens (1999) and Nevo (2003) used estimated selection probabilities to compute the parameters of interest (the linear regression or GMM model respectively) without focusing on selection probabilities by themselves.

I use the MN sampling literature approach to reformulate my problem so that I can directly use Hellerstein and Imbens (1999)'s and Nevo (2003)'s methodology. I propose a GMM type estimation procedure based on Nevo (2003) which, in contrast to Lancaster and Imbens (1996), only requires knowledge of a finite set of moments of covariates. I extend Nevo (2003)'s procedure by addressing asymmetries in data availability. In many cases, for example in rare events studies or in marketing research, there is more complete information for a given choice outcome.[2] In the opposite case, full data on covariates is only available for the whole population, while information for particular choices is private and unobservable.[3] My method allows that either the population or the choice-based distribution of covariates to be replaced by only finite information (such as several moments or quantiles). Most importantly, for some cases, information on some regressors can be completely missing from, for example, population data. This paper extends Nevo (2003)'s approach to the case when the refreshment sample is smaller than the primary one. The researcher just needs to treat the additional sample as primary and the primary sample as auxiliary, estimate the selection probabilities using my approach, and use these probabilities to correct for non-random selection in panel models with non-random attrition.

As an empirical application of the proposed procedure, I estimate the probability of a two-car accident based on the data on all police reported accidents in Seattle from 2002-2011. The results suggest that the drinking status of drivers has no significant effect on probability of a two-car accident.

This paper is organized as follows. Section 2 formally defines the underlying data generating process and the available data structures. Section 3 presents the estimator. I show the empirical application in Section 4. Section 5 concludes.

[2] For example, ecological studies, habitat studies of so-called presence-only data, e.g. Graham et al. (2004), Pearce and Boyce (2005)

[3] For example, different types of elections.

3.1. Underlying data generating process and data structure

3.1.1. Data generating process

Let $\mathbf{y}$ be binary and $\mathbf{x}$ be a vector of attributes with support $X \subseteq \mathbb{R}^d$. Assume that $\mathbf{x}$ has an unknown c.d.f F_x and that

$$\Pr(\mathbf{y} = 1 | \mathbf{x} = x) = G(x, \theta_0),$$

where $\theta_0 \in \Theta \subseteq \mathbb{R}^d$ is a vector of parameters to be estimated, and G is a known function. Let $q = \Pr(\mathbf{y} = 1) = \mathbb{E}[\, G(\mathbf{x}, \theta_0)\,] > 0$ and let $\mathbf{z}$ be a random variable such that $\Pr(\mathbf{z} \leq t) = \Pr(\mathbf{x} \leq t | \mathbf{y} = 1)$ for all t. I have introduced $\mathbf{z}$ in order to make distinction between a sample from the population and a contaminated one. If the econometrician observes realizations of $\mathbf{z}$ then she has a contaminated sample.

3.1.2. Data structure

There are different possible data sets the econometrician could obtain: (i) The econometrician could observe realizations of $\mathbf{z}$ only. Moreover, she could know some information about the marginal distribution of $\mathbf{x}$, for example, she could observe $\mathbb{E}[\, h(\mathbf{x})\,] = \bar{h}_x$ for a known function $h_x : X \to \mathbb{R}^d$. I focus on the exact identification case, although the number of available moments can be greater than dimension of $\mathbf{x}$. (ii) Alternatively, the researcher could obtain a random sample of $\mathbf{x}$ and know some moments of $\mathbf{z}$. That is, she could observe $\mathbb{E}[\, h_z(\mathbf{z})\,] = \bar{h}_z$ for a known function h_z. The procedure allows both possibilities. Importantly, there is no big difference between the above data structure types. The researcher can estimate the parameters of interest whatever data she has.

3.2. Identification and estimation with known q

The estimation strategy is easy to motivate. First, assume for a moment that $\mathbf{x}$ is a continuously distributed random variable with p.d.f. f_x. From Bayes' Rule

$$f_{x|y=1}(x) = \frac{\Pr(\mathbf{y}=1|\mathbf{x}=x)}{\Pr(\mathbf{y}=1)} f_x(x) = \frac{G(x,\theta_0)}{q} f_x(x)$$

Hence,

$$\bar{h}_x = \mathbb{E}[\,h_x(x)\,] = q\mathbb{E}\left[\frac{h_x(\mathbf{x})}{G(\mathbf{x},\theta_0)}\bigg|\mathbf{y}=1\right] = q\mathbb{E}\left[\frac{h_x(\mathbf{z})}{G(\mathbf{z},\theta_0)}\right]$$

Generally, the following claim allows me to identify and estimate the parameters of interest θ_0 using GMM.

Claim 3.1 *For a given function function $h : X\times\Theta \to \mathbb{R}^m$, if (i) $\Pr(\Pr(\mathbf{y}=1|\mathbf{x})>0) = 1$, and (ii) for all $\theta\in\Theta$ $\mathbb{E}\left[\left|\frac{h(\mathbf{x},\theta)}{\Pr(\mathbf{y}=1|\mathbf{x})}\right|\right] < \infty$, then for all $\theta\in\Theta$*

$$\mathbb{E}[\,h(\mathbf{x},\theta)\,] = \Pr(\mathbf{y}=1)\mathbb{E}\left[\frac{h(\mathbf{x},\theta)}{\Pr(\mathbf{y}=1|\mathbf{x})}\bigg|\mathbf{y}=1\right]$$

Notice that Claim 1 allows me to construct a system of moments both for $\mathbf{x}$ and for $\mathbf{z}$. Indeed, if instead of h I apply Claim 1 to $\tilde{h}(x,\theta) = G(x,\theta)h(x,\theta)/q$, I can write

$$\mathbb{E}[\,h(\mathbf{x},\theta)|\mathbf{y}=1\,] = \mathbb{E}\left[\frac{G(\mathbf{x},\theta)h(\mathbf{x},\theta)}{q}\right]$$

Denote

$$m_z(z,\theta) \equiv q\frac{h_x(z)}{G(z,\theta)} - \bar{h}_x, \qquad m_x(x,\theta) \equiv h_z(x)G(x,\theta) - \bar{h}_z q$$

Then I have the following systems of unconditional moments of $\mathbf{x}$ and $\mathbf{z}$:

$$\mathbb{E}[\,m_x(\mathbf{x},\theta_0)\,] = 0 \quad (3.1)$$

$$\mathbb{E}[\,m_z(\mathbf{z},\theta_0)\,] = 0 \quad (3.2)$$

The identification and estimation problem then is simply a question of the uniqueness of the solution to the system of equations (1) or (2). It follows that there is no difference between cases (i) and (ii) (See section 2.2). If you have sample of $\mathbf{z}$ and know some moments of $\mathbf{x}$, you can construct system (1) and, if you can obtain sample from $\mathbf{x}$ and know moments of $\mathbf{z}$, you can estimate the parameters by using (2).

3.2.1. Identification

In the estimation part I assume that the parameters of interest are point identified. In this subsection I provide sufficient conditions for identification in some particular specifications of the model. As a sufficient condition for point identification I use a standard rank condition. Let $\xi \in \{x, z\}$, then

Proposition 3.2 *If for all $\theta \in \Theta$, $\mathbb{E}[\,\partial_{\theta^T} m_\xi(\xi, \theta)\,]$ has full column rank then θ_0 is identified.*

Generally, the rank condition is hard to verify. However, for the special case which is used in the empirical application, I can derive a simple sufficient conditions.

Proposition 3.3 *Let $h(\xi) = \xi$ and $G(\xi, \theta) = F(\xi^T\theta)$. If (i) Support of ξ is compact; (ii) $F(\cdot)$ is a strictly positive and continuously differentiable strictly increasing (decreasing) function; (iii) For all $\theta \in \Theta$, $\mathbb{E}[\,\partial_{\theta^T} m_\xi(\xi, \theta)\,]$ exists; and (iv) $\mathbb{E}\left[\xi\xi^T\right] > 0$;then θ_0 is identified.*

Proof. See appendix. ■

Proposition 2 has a direct application to the Probit and the Logit models. Consider the following model

$$\begin{cases} \mathbf{y} = \mathbb{1}(\mathbf{x}^T\theta_0 + \epsilon \geq 0) \\ \epsilon \sim F(\epsilon) \\ \epsilon \text{ is independent of } \mathbf{x}, \end{cases}$$

where F is the c.d.f. of the standard normal or the logistic distribution. Then $G(x, \theta) = 1 - F(-x^T\theta)$ and conditions of Proposition 2 are satisfied.

Proposition 2 is similar to Proposition 1 in Nevo (2003).[4] Proposition 2 does require knowing moments for each of the covariates. However, in some cases, if the researcher cannot obtain a moment for one of them, she can use different moments

[4]Nevo does not allow for the presence of an intercept and considers the case when $\xi = z$. Since in my setting q is known, I still can identify the intercept.

of the rest of covariates to identify the parameter of interest. For instance, consider the following example.

Example 3.1 Let

$$G(z, \theta_0) = \frac{1}{1 + \exp\{\theta_{00} + \theta_{01} z_1 + \theta_{02} z_2\}}$$

$z = (z_1, z_2) \in \{-1, 0, 1\} \times \{0, 1\}$ is distributed according to

$$\begin{cases} \Pr(z_1 = 1, z_2 = 1) = 2/3 \\ \Pr(z_1 = 0, z_2 = 0) = \Pr(z_1 = -1, z_2 = 0) = 1/6 \\ \Pr(z_1 = 0, z_2 = 1) = \Pr(z_1 = -1, z_2 = 1) = \Pr(z_1 = 1, z_2 = 0) = 0, \end{cases}$$

and

$$h(z) = \begin{bmatrix} 1 \\ z_1 \\ z_1^2 \end{bmatrix}$$

That is, instead of using a moment of $\mathbf{z}_2$, I use the second moment of $\mathbf{z}_1$ to identify θ_{02}. Then

$$\begin{aligned} -\mathbb{E}[\,\partial_{\theta^T} m(\mathbf{z}, \theta)\,]/q &= \mathbb{E}\left[\exp\{\theta_0 + \theta_1 z_1 + \theta_2 z_2\} \begin{bmatrix} 1 & z_1 & z_2 \\ z_1 & z_1^2 & z_1 z_2 \\ z_1^2 & z_1^3 & z_1^2 z_2 \end{bmatrix} \right] \\ &= e^{\theta_0} \left\{ 2/3 e^{\theta_1 + \theta_2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + 1/6 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + 1/6 e^{-\theta_1} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix} \right\} \\ &= e^{\theta_0}/6 \begin{bmatrix} 4e^{\theta_1+\theta_2} + 1 + e^{-\theta_1} & 4e^{\theta_1+\theta_2} - e^{-\theta_1} & 4e^{\theta_1+\theta_2} \\ 4e^{\theta_1+\theta_2} - e^{-\theta_1} & 4e^{\theta_1+\theta_2} + e^{-\theta_1} & 4e^{\theta_1+\theta_2} \\ 4e^{\theta_1+\theta_2} + e^{-\theta_1} & 4e^{\theta_1+\theta_2} - e^{-\theta_1} & 4e^{\theta_1+\theta_2} \end{bmatrix} \end{aligned}$$

and the right hand side matrix has the same rank as

$$\begin{bmatrix} 4e^{\theta_1+\theta_2} + 1 + e^{-\theta_1} & 4e^{\theta_1+\theta_2} - e^{-\theta_1} & 4e^{\theta_1+\theta_2} \\ -2e^{-\theta_1} - 1 & 2e^{-\theta_1} & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

which has full rank for all finite θ. Hence, θ_0 is identified.

$h(\mathbf{z})$ can be treated as an instrument for $\mathbf{z}$. If $h(\mathbf{z})$ is a "good" instrument, then θ_0 might still be identified. In the above example, using the fact that $\mathbf{z}_1$ and $\mathbf{z}_2$ are not independent, $\mathbf{z}_1^2$ is used as an instrument for $\mathbf{z}_2$.

3.2.2. Estimation

In this section I assume that $\xi = z$. The asymptotic properties for the case $\xi = x$ can be derived by analogy. I use the standard two-step GMM to estimate the parameters of interest. Denote

$$\hat{m}(\theta) = 1/n \sum_{i=1}^{n} m_z(z_i, \theta), \qquad \tilde{V} = 1/n \sum_{i=1}^{n} m_z(z_i, \tilde{\theta}) m_z^T(z_i, \tilde{\theta})$$

$$\tilde{\theta} = \arg\min_{\theta \in \Theta} [\hat{m}^T(\theta)\hat{m}(\theta)], \qquad \hat{\theta} = \arg\min_{\theta \in \Theta} [\hat{m}^T(\theta)\tilde{V}^{-1}\hat{m}(\theta)]$$

Then under standard regularity conditions listed below, my GMM estimator is consistent and asymptotically normal.

Theorem 3.4 *If (i) Θ and X are compact; (ii) $G(x, \theta)$ and its derivative with respect to θ are continuous and bounded away from zero for all $\theta \in \Theta$ and $x \in X$; (iii) θ_0 is the unique solution to $\mathbb{E}[\, m_z(\mathbf{z}, \theta)\,] = 0$; (iv) The matrix $V = \mathbb{E}\Big[\, q^2 h(\mathbf{z}) h^T(\mathbf{z}) / G^2(\mathbf{z}, \theta_0)\,\Big] - \bar{h}\bar{h}^T$ is nonsingular; and (iv) The matrix $A = \mathbb{E}[\, q h(\mathbf{z}) \partial_{\theta^T} G(\mathbf{z}, \theta_0) / G^2(\mathbf{z}, \theta_0)\,]$ is of full column rank; then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, (A^T V^{-1} A)^{-1})$$

Proof. Consistency and asymptotic normality of the estimator can be proved as in Newey and McFadden (1994). ■

In the above analysis I assume that $\bar{h}$ is known exactly. The next subsection generalizes the result to the case when $\bar{h}$ is not known for sure.

3.2.3. Accounting for sampling error in the moment restrictions

Assume that, instead of knowing $\bar{h} = \mathbb{E}[\, h(\mathbf{x})\,]$, we observe an independent sample $\{x_i\}_{i=1}^{m}$ along with the primary sample $\{z_i\}_{i=1}^{n}$. Hence, $\hat{h} = 1/m \sum_i h(x_i)$ is a consistent estimator of $\bar{h}$. This estimate is asymptotically normal with a variance $\mathbb{E}\left[\,h(\mathbf{x})h(\mathbf{x})^T\,\right]$. If I treat $\hat{h}$ as $\bar{h}$ then, as shown in Hellerstein and Imbens (1999), I can still estimate θ_0. However, I need to take into account that the samples might converge to infinity with different rates. If m/n converges to zero, then the sampling error in $\hat{h}$ can be neglected. If m/n converges to infinity, then I can treat $\{x_i\}_{i=1}^m$ as the primary sample and $\{z\}_{i=1}^m$ as auxiliary, and use the above procedure to the system of moments (1).[5] Hence, even if the axillary sample is smaller than the primary one, the researcher can estimate the conditional choice probabilities by conducting the procedure as if the auxiliary sample was primary. If m/n converges to an integer $k \neq 0$ then, following Hellerstein and Imbens (1999), I can describe the asymptotic behavior as follows. Suppose we have n observations of $\mathbf{t}$, where t_i consists of $(z_i, h_{i1}, h_{i2}, \ldots, h_{ik})$. Denote

$$\hat{m}(\theta) = 1/n \sum_{i=1}^{n} \left[\frac{qh(z_i)}{G(z_i,\theta)} - 1/k \sum_{l=1}^{k} h_{il} \right], \qquad \tilde{V} = 1/n \sum_{i=1}^{n} m_z(z_i,\tilde{\theta}) m_z^T(z_i,\tilde{\theta})$$

$$\tilde{\theta} = \arg\min_{\theta\in\Theta} [\hat{m}^T(\theta)\hat{m}(\theta)], \qquad \hat{\theta} = \arg\min_{\theta\in\Theta} [\hat{m}^T(\theta)\tilde{V}^{-1}\hat{m}(\theta)]$$

Similarly to Theorem 1 the following theorem describes the large sample properties of the estimator.

Theorem 3.5 *If (i) Θ and X are compact; (ii) $G(x,\theta)$ and its derivative with respect to θ are continuous and bounded away from zero for all $\theta \in \Theta$ and $x \in X$; (iii) θ_0 is the unique solution to $\mathbb{E}[\,m(\mathbf{z},\theta)\,] = 0$; (iv) The matrix $V = \mathbb{E}\left[\,q^2 h(\mathbf{z})h^T(\mathbf{z})/G^2(\mathbf{z},\theta_0)\,\right] - 1/k^2 \sum_{l=1}^{k} \mathbb{E}[\,h_{1l}\,] \sum_{l=1}^{k} \mathbb{E}\left[\,h_{1l}^T\,\right]$ is nonsingular; (iv) The matrix $A = \mathbb{E}[\,qh(\mathbf{z})\partial_{\theta^T} G(\mathbf{z},\theta_0)/G^2(\mathbf{z},\theta_0)\,]$ is of full column rank; then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, (A^T V^{-1} A)^{-1})$$

Proof. The same as the proof of Theorem 1. ∎

[5] This case is not considered in Hellerstein and Imbens (1999). Nevo (2003) mentions that in this case his procedure does not work.

3.3. Empirical application

In this section, I use the procedure described above to estimate the conditional probability of a two-car accident in Seattle. Since I observe only car interactions that resulted in an accident, I have a contaminated sample. Moreover, the sobriety level of drivers, an important characteristic that affects the probability of collision, is observable only in the collision data. Thus, there is no auxiliary information on the populational drinking habits. However, I still can estimate the effect of alcohol consumption. To do this, I use information on the age distribution of the drivers, because age and drinking habits are correlated. First, I start by describing the data set. I proceed by presenting a simple model of collisions. Finally, I continue with the estimation results.

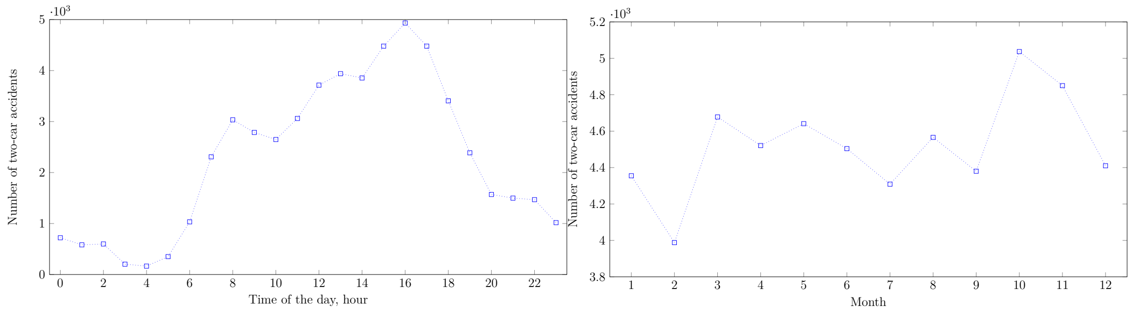
3.3.1. Data on collisions

The data on two-car accidents in Seattle is provided by the Statewide Travel and Collision Data Office (STCDO) of the Washington State Department of Transportation (WSDOT). The data set contains information on all police reported car accidents in Seattle from 2002 to 2011.[6] Among the variables collected in each accident are information on the time, location of the crash, weather conditions and road surface condition at the moment of the accident; as well as characteristics of drivers involved such as age, sex and whether the drivers were under the influence of alcohol. I construct dummy variables for each hour of the day (e.g. 12am-1am), each month of the year, weekend (1 if it is Friday, Saturday or Sunday), month, gender (1 if male), road condition (1 if the road surface is not dry) and the sobriety level of the drivers.[7] If the road condition is unknown then I use the value of the weather conditions variable (rain, snow or ice). I exclude from the sample any crash in which either information on weather and road conditions or characteristics of one or more of the drivers is missing. I use 54171 observations.

Table 3.1, Figures 3.1-3.3 present summary statistics for the data in the sample. Drivers in two-car crashes are mostly male. Roughly 5 percent of drivers had been drinking. Somewhat less than half of the accidents happened during weekends. The

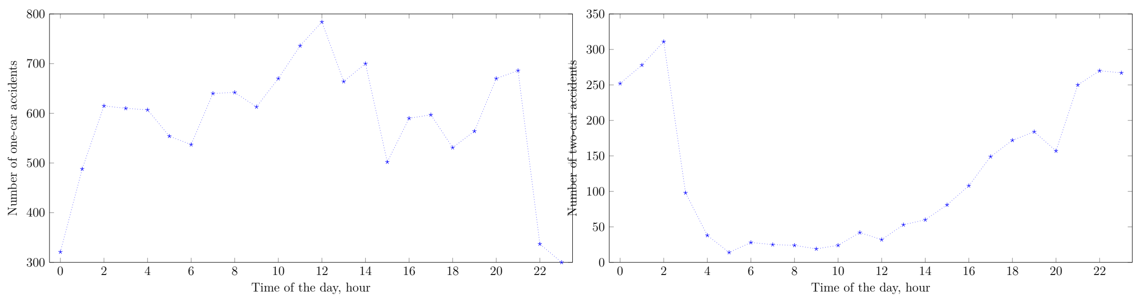
[6]Actually, I have data on collisions in 2001. Unfortunately the data was not collected properly and most of the observations are considered to be missing.

[7]Sobriety level of each driver is either "Had been drinking-ability impaired", "Had been drinking-ability not impaired" or "Had not been drinking". I treat driver as sober if she had not been drinking.



(a) by hour of the day **(b)** by month of the year

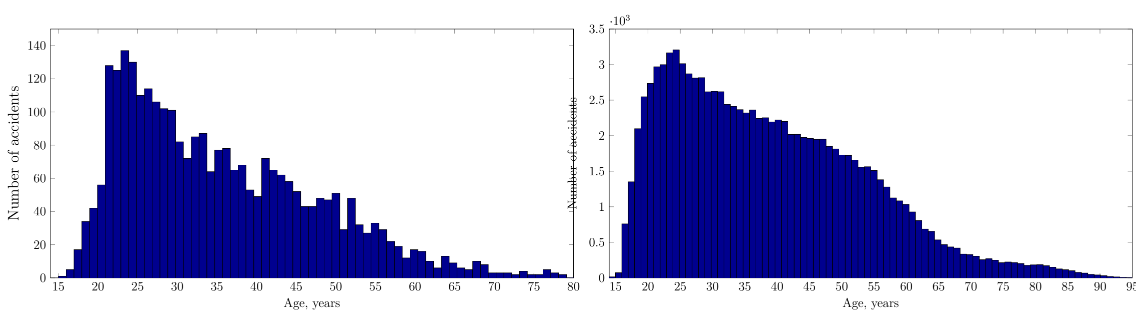
Figure 3.1 – Number of two-car accidents



(a) one-car **(b)** 'drunk' two-car

Figure 3.2 – Number of accidents by hour of the day

distribution of two-car accidents by hour of a day looks similar to the traffic distribution during a day (Figure 3.4a). The fraction of two-car accidents between 2am and 6am is small. However, the fraction of one-car accidents during the same time interval is much higher (Figure 3.2a). Figure 3.2b represents the distribution of the number of accidents where at least one driver was drunk by the hour of the day. Most of this crashes happened during the night. Although there are drivers who had been drinking between 6am and 4pm. Figures 3.3a and 3.3b present the age distribution



(a) 'drunk' **(b)** 'sober'

Figure 3.3 – Number of two-car accidents by age of drivers

of drivers in the sample. There is a sudden increase in the number of accidents for 21 years old drivers.

Table 3.1 – Summary Statistics for Two-car Accident in the Sample, 2001-2011

Variable	Mean
Total number of two-car accidents	54,171
Percentage of all accidents in two-car accidents	
Rain or snow	21.0
Weekend	40.8
Age of drivers	39
Percentage of drivers in two-car accident	
Male	61
Two drinking drivers	0.3
One drinking, one sober driver	4.9
Two sober drivers	94.8

3.3.2. Model of a two-car collision

My model of two car accidents is based on the one presented in Levitt and Porter (2001). Assume that there is a stream of two-cars interactions on the road. Each interaction is characterized by a pair of random variables $(\mathbf{y}, \mathbf{x})$. The vector of attributes $\mathbf{x}$ consists of binary characteristics of the environment $\mathbf{x}_e$: road condition, time of the day, weekend day, month of the year and an intercept; and of the drivers involved $\mathbf{x}_1$ and $\mathbf{x}_2$: age, gender and sobriety level. Assume that, conditional on environment, the characteristics if drivers involved in an interaction are i.i.d.. $\mathbf{y}$ represents the outcome of the interaction ($\mathbf{y} = 1$ if there is an accident).

Let $F(x_e^T\theta_{00} + x_i^T\theta_{10})$ be the probability that a driver i with characteristics (x_e, x_i) makes a mistake during an interaction (F is the c.d.f. of the standard normal or the logistic distribution). Also assume that a crash results from a single driver's mistake. This assumption excludes cases when mistakes of both drivers have caused the crash. Then

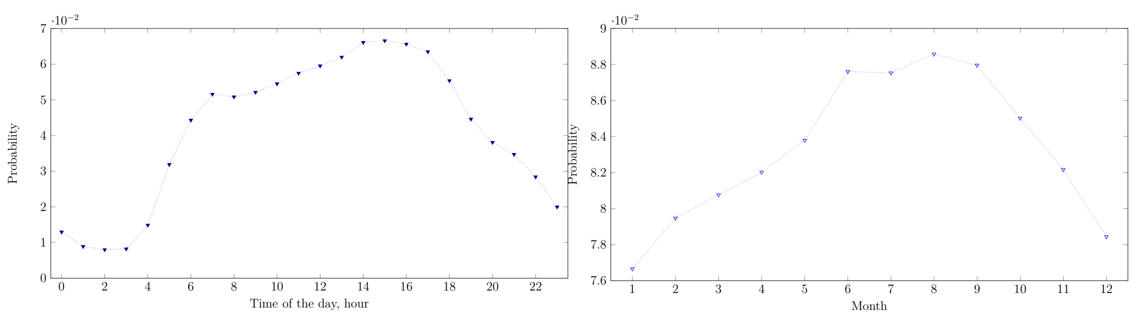
$$\Pr(\mathbf{y} = 1|x) = G(x, \theta_0) \equiv F(x_e^T\theta_{00}+x_1^T\theta_{10})+F(x_e^T\theta_{00}+x_2^T\theta_{10})-F(x_e^T\theta_{00}+x_1^T\theta_{10})F(x_e^T\theta_{00}+x_2^T\theta_{10})$$

3.3.3. Estimating $\bar{h}$ and q

In order to construct system of moments, I need to choose a function h. From the auxiliary data, I can estimate moments of all of the covariates except for sobriety level. There is no reliable information on the drinking habits of drivers. Since all of the covariates except age are binary, I have to use some nonlinear function of age as an additional moment to identify the coefficient corresponding to sobriety level. From the auxiliary data, I know the distribution of age over 22 groups for both male and female drivers. In order to use all available information, I add $\mathbb{1}(\text{age} \in \text{AgeGroup}_i)$ and $\mathbb{1}(\text{age} \in \text{AgeGroup}_i)\mathbb{1}(\text{gender} = \text{male})$ for $i = 1, \ldots, 21$, as additional moments. As a result, I end up with 80 moments and 41 parameters to be estimated.

To apply the estimation procedure described above I need to know $\mathbb{E}[\,h(x)\,]$. I assume that the distribution of drivers on the road is equal to distribution of registered drivers. This assumption might not be valid. It is possible that males on average drive more than females.

The age and sex distributions of drivers in Seattle were obtained from the website of the U.S. Department of Transportation.[8] Since the probabilities of a two-car interaction and bad weather conditions change during the day and the year, estimation of moments of time and weather related variables requires more attention. To estimate the probability of a two-car interaction in a given time interval I use data on traffic volumes from 30 locations from 2004-2009 in Seattle from STCDO of WSDOT.[9] The



(a) by hour of the day **(b)** by month of the year

Figure 3.4 – Traffic distribution

data set allows the researcher to compute the traffic volume at a given location, year and day within 5 minute intervals. I use the fraction of the aggregate traffic volume in a given time interval (e.g. 10am-11am, weekend) to the total traffic volume as a

[8] http://www.fhwa.dot.gov/policyinformation/quickfinddata/qfdrivers.cfm

[9] I could not obtain traffic data for the whole city due to the size of data set. The data set I have takes roughly 30 Gb of space and was send to me by regular mail on 10 DVD discs.

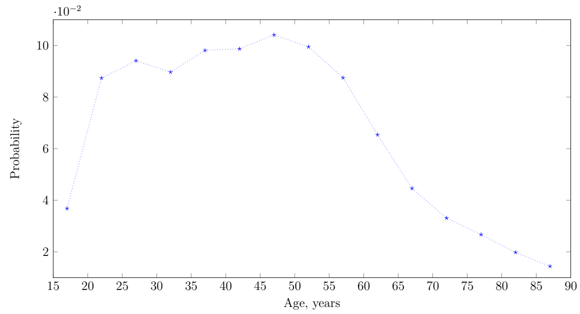


Figure 3.5 – Age distribution

proxy for the expectations of the corresponding time related variables. As a proxy for q I use the fraction of two-car accidents to the total volume of traffic in Seattle.[10] The probability of having rainy weather was calculated based on information supplied by the National Weather Service.[11] I calculated the probabilities of the precipitation for each day of the year. Then I adjusted this probabilities by the traffic distribution over the year. Data on moments used in estimation are presented in Table 3.2 and Figures 3.4-3.5.

Table 3.2 – Moments used in estimation

h_i	wheather	weekend	age/100
$\mathbb{E}[h_i]$	0.1353	0.4170	0.4515

3.3.4. Estimation results

Estimation results are presented in Table 3.3. Most of the coefficients are significantly different from 0. The probability of having a two-car crash between 3am and 7am is smaller than the probability of an accident during the rest of the day. Driving during the winter is safer than driving during the summer. These results might be explained by difference in traffic volumes. Men are much worse drivers than women. The gender effect is of the same magnitude as the effect of bad weather. The probability of making a mistake on the road is decreasing with age. Surprisingly, the effect of being drunk is not significantly different from 0, although it is positive. There are several possible explanations of this result. First, I considered drivers that 'had been

[10] WSDOT reports average daily traffic, e.g. http://www.seattle.gov/transportation/docs/2009TrafficReport.pdf.

[11] http://www.komonews.com/weather/faq/4308877.html

drinking - ability not impaired' as drunk drivers.[12] Second, the data set includes only police-reported crashes. Minor accidents that involved drunk drivers might not be in the sample. Finally, sobriety level is correlated with the time of the day, hence, there might be lack of identification.

Table 3.3 – Estimation results

	Probit		Logit	
Variable	θ	s.e.	θ	s.e.
Constant	-4.3306***	(0.0176)	-12.0001***	(0.0793)
12am-1am	-0.0157	(0.0159)	0.0000	(0.0689)
1am-2am	0.0000	(0.0243)	0.1381	(0.0993)
2am-3am	0.0000	(0.0301)	0.2661**	(0.1205)
3am-4am	-0.2254***	(0.0306)	-0.9030***	(0.1299)
4am-5am	-0.2839***	(0.0250)	-1.2309***	(0.1086)
5am-6am	-0.2779***	(0.0244)	-1.3462***	(0.1093)
6am-7am	-0.0816***	(0.0222)	-0.4298***	(0.0994)
7am-8am	0.0548**	(0.0216)	0.1918**	(0.0976)
8am-9am	0.1469***	(0.0217)	0.6064***	(0.0983)
9am-10am	0.1393***	(0.0220)	0.5738***	(0.0996)
10am-11am	0.1417***	(0.0219)	0.5815***	(0.0990)
11am-12pm	0.1680***	(0.0219)	0.7075***	(0.0983)
12pm-1pm	0.1895***	(0.0219)	0.8051***	(0.0991)
1pm-2pm	0.2008***	(0.0212)	0.8592***	(0.0956)
2pm-3pm	0.1613***	(0.0213)	0.6749***	(0.0962)
3pm-4pm	0.1949***	(0.0209)	0.8297***	(0.0947)
4pm-5pm	0.2016***	(0.0208)	0.8571***	(0.0939)
5pm-6pm	0.1774***	(0.0198)	0.7549***	(0.0891)
6pm-7pm	0.1446***	(0.0191)	0.6139***	(0.0864)
7pm-8pm	0.0777***	(0.0175)	0.3218***	(0.0791)
8pm-9pm	0.0032	(0.0166)	0.0000	(0.0723)
9pm-10pm	0.0000	(0.0134)	0.0000	(0.0585)

Notes: *** indicates statistically significantly different from zero at the 1% level; ** indicates 5% level; * indicates 10% level.

[12] I ran the procedure when this type of drivers were considered to be sober. Unfortunately, the number of drunk accidents became very small, so I couldn't identify the effect of drinking.

Table 3.3 – Continued

Variable	Probit θ	Probit s.e.	Logit θ	Logit s.e.
10pm-11pm	0.0245*	(0.0130)	0.1247**	(0.0575)
January	-0.0040	(0.0080)	0.0000	(0.0349)
February	0.0089	(0.0077)	0.0432	(0.0335)
March	0.0344***	(0.0076)	0.1657***	(0.0333)
April	0.0365***	(0.0076)	0.1744***	(0.0334)
May	0.0448***	(0.0075)	0.2156***	(0.0329)
June	0.0390***	(0.0080)	0.1849***	(0.0345)
July	0.0498***	(0.0076)	0.2313***	(0.0334)
August	0.0469***	(0.0075)	0.2198***	(0.0329)
September	0.0376***	(0.0078)	0.1802***	(0.0340)
October	0.0387***	(0.0074)	0.1805***	(0.0327)
November	0.0037	(0.0078)	0.0280	(0.0341)
Weather	0.2464***	(0.0035)	1.1005***	(0.0155)
Weekend	-0.0337***	(0.0034)	-0.1464***	(0.0152)
Age	-0.2459***	(0.0838)	0.0000***	(0.3743)
Gender	0.1449***	(0.0040)	0.6652***	(0.0181)
Sobriety level	0.2560	(0.2044)	0.2344	(0.6786)
Age^2	-1.0531***	(0.1017)	-5.9710***	(0.4568)

Notes: *** indicates statistically significantly different from zero at the 1% level; ** indicates 5% level; * indicates 10% level.

3.4. Summary

This paper proposes a method to estimate binary models with a pure-choice based data or a data with unobserved responses in the presence of additional information that comes in the form of the finite set of moments. Importantly, the pure-choice based data problem is dual to the problem of data with unobserved response. Hence, the procedure can be used in estimation of inverse probability weights in data sets with non-random attrition even if the refreshment sample is much smaller than the primary sample. I applied the procedure to estimation of the probability of a two-car accident in Seattle.

Appendix A

Omitted proofs for Chapter 1

A.1. Proof of Proposition 1.4

Let (β, h) satisfy assumptions 6 and 7, and suppose that $\beta \neq \beta_0$. We will show that there exists a set $X' \subseteq X$ such that $\Pr(\mathbf{x} \in X') > 0$ and $\mu_0(x) \neq \mathbb{E}[\, h(\mathbf{e}, x)|x\,]$ for all $x \in X'$. Consequently, h is not consistent with the observed data, and (β, h) does not belong to the sharp identified set.

By Assumption 7, there exists some i, $y = (y_i, y_{-i})$ and X' for which either (1.7) or (1.8) holds. Without loss of generality suppose that (1.7) holds (the alternative case is analogous). Let $\mathbf{x}_j$ be the covariate from Assumption 6 corresponding to y_{-i}. Also, consider an arbitrary sequence $(\xi^k)_{k\in\mathbb{N}}$ diverging to infinity, and let (X^k) be the sequence of sets given by $X^k = \left\{x \in X' \mid x_j \geq \xi_j^k\right\}$.

Note that:

$$\begin{aligned}
\mathbb{E}\left[\, \mathbf{h}(y_i)|X^k \,\right] &\leq \mathbb{E}\left[\, \mathbf{h}(\mathbf{BR}_i(y_{-i};\beta)) \cdot \Pr(y_i \in \mathbf{BR}_i(y_{-i};\beta))|X^k \,\right] \\
&\quad \ldots + \mathbb{E}\left[\, \mathbf{h}(Y \setminus \mathbf{BR}_i(y_{-i};\beta))|X^k \,\right] \\
&\leq \Pr(y_i \in \mathbf{BR}_i(y_{-i};\beta)|X') + \mathbb{E}\left[\, 1 - \mathbf{h}(\mathbf{BR}_i(y_{-i};\beta))|X^k \,\right],
\end{aligned}$$

where the second equality uses the fact that $\mathbf{u}_i$ and $\mathbf{x}_j$ are independent conditional on $\mathbf{x}_{-j}$. Condition (1.6) implies that the second term vanishes in the limit and thus:

$$\lim_{k\to\infty} \mathbb{E}\left[\, \mathbf{h}(y_i)|X^k \,\right] \leq \Pr\left(y_i \in \mathbf{BR}_i(y_{-i};\beta) \,\middle|\, X'\right). \tag{A.1}$$

On the other hand, for the true data generating process we have that:

$$\mathbb{E}\left[\, \mathbf{h}_0(y_i)|X^k \,\right] \geq \Pr\left(\mathbf{y}_i \in \mathbf{BR}_i(y_{-i};\beta_0) \,\wedge\, \mathbf{BR}_i(y_{-i};\beta_0) = \{y_i\} \,\middle|\, X^k\right)$$

$$= \mathbb{E}\Big[\mathbf{h}_0(\mathbf{BR}_i(y_{-i};\beta_0)) \,\Big|\, X^k, \mathbf{BR}_i(y_{-i};\beta_0) = \{y_i\} \Big]$$
$$\ldots \times \Pr\Big(\mathbf{BR}_i(y_{-i};\beta_0) = \{y_i\} \,\Big|\, X'\Big)$$

Condition (1.7) implies that $\Pr\Big(\mathbf{BR}_i(y_{-i};\beta_0) = \{y_i\} \,\Big|\, X'\Big) > 0$. Therefore, condition (1.6) implies that

$$\lim_{k\to\infty} \mathbb{E}\Big[\mathbf{h}_0(\mathbf{BR}_i(y_{-i};\beta_0)) \,\Big|\, X^k, \mathbf{BR}_i(y_{-i};\beta_0) = \{y_i\} \Big] = 1,$$

and, consequently,

$$\lim_{k\to\infty} \mathbb{E}\Big[[\mu_0(\mathbf{x})](y_i) | X^k \Big] = \lim_{k\to\infty} \mathbb{E}\Big[\mathbf{h}_0(y_i) | X^k \Big] \geq \Pr\Big(\mathbf{BR}_i(y_{-i};\beta_0) = \{y_i\} \,\Big|\, X'\Big). \tag{A.2}$$

From (A.1) and (A.2), and the fact that h_0 is consistent with the observed data, it follows that:

$$\lim_{k\to\infty} \mathbb{E}\Big[[\mu_0(\mathbf{x})](y_i) - \mathbf{h}(y_i) | X^k \Big] \geq \Pr\Big(\mathbf{BR}_i(y_{-i};\beta_0) = \{y_i\} \,\Big|\, X'\Big)$$
$$\ldots \quad - \Pr\Big(y_i \in \mathbf{BR}_i(y_{-i};\beta) \,\Big|\, X'\Big) > 0$$

Since $\Pr(X^k) > 0$ for all k, it follows that h is not consistent with the data. ■

A.2. Omitted details from section 1.6.5

A.2.1. BNE of the incomplete information game

First we need to characterize the set of BNE. Note that, interim preferences over pure strategies are represented by the expected utility function:

$$v_i(a; x, e_i) = \Big[\beta_{0i} x_i - e_i - \beta_{03} \Pr(a_{-i}(\mathbf{e}_{-i}) = 1 \,|\, x)\Big] a_i(e_i). \tag{A.3}$$

Since (A.3) is linear in a_i, and the coefficient is monotone in e_i, it follows that best response correspondence is monotone in e_i and the agent is generically not indifferent. Hence, there can only be pure strategy BNE and they need to be threshold strategies as in (1.12). Furthermore, such strategies constitute a BNE if and only if each agent

i is indifferent between entering and not entering, which is exactly (1.13).

To establish existence, note that (1.13) can be rewritten as:

$$\bar{e}_1 = f_1(\bar{e}_2) \equiv \beta_{01}x_1 - \beta_{03}\Phi(\bar{e}_2) \qquad \wedge \qquad \bar{e}_1 = f_2(\bar{e}_2) \equiv \Phi^{-1}\left(\frac{\beta_{02}x_2 - \bar{e}_2}{\beta_{03}}\right),$$

and let $f_3(e_2) = f_1(e_2) - f_2(e_2)$, for $e_2 \in (\beta_{02}x_2 - \beta_{03}, \beta_{02}x_2)$. Note that f_3 is continuous, and converges to $-\infty$ and ∞ at the extremes of these interval. Hence, the mean value theorem implies that there exist some e_2 such that $f_3(e_2) = 0$, i.e., $f_1(e_2) = f_2(e_2)$. There exists a set of parameters for which there are multiple equilibria, but all equilibria are of this form.

A.2.2. Proof of Proposition 1.7

Note that $x(t)$ makes the incomplete information game symmetric, so that $\bar{e}_1(x(t)) = \bar{e}_2(x(t))$ for all t. Let $\hat{e}(t) \in (\beta_{03}(t-1), \beta_{03}t)$ be the minimum solution to (1.13), so that the probability of $(1,1)$ in any BNE is bounded below by the probability of of $(1,1)$ in the BNE with thresholds $\hat{e}(t)$, i.e.,

$$m_{\text{BNE}}(t) = \Phi^2(\hat{e}(t)). \tag{A.4}$$

On the other hand, for the complete information game, there are three regions of realizations of e_i. If $e_i > \beta_{03}t$ then not entering is dominant for i, if $e_i < \beta_{03}(t-1)$ then entering is dominant for i, and, if $e_i \in (\beta_{03}(t-1), \beta_{03}t)$ then i wants to enter if $-i$ is not entering, and wants to stay out if $-i$ is entering. Hence, the outcome $(1,1)$ can occur in equilibrium only if $e_i < \beta_{03}(t-1)$ for $i = 1,2$ or if $e_i \in (\beta_{03}(t-1), \beta_{03}t)$ for $i = 1,2$. In the first case, $(1,1)$ is the only Nash equilibrium. In the former case, the game has two pure equilibria $(0,1)$ and $(1,0)$, and a strictly mixed equilibrium in which $(1,1)$ occurs with some probability $q \in (0,1)$. This implies that

$$\begin{aligned} M_{\text{NE}}(t) &= \Phi^2(\beta_{03}(t-1)) + \Big[\Phi^2(\beta_{03}t) - \Phi(\beta_{03}(t-1))\Big]^2 q^2 \\ &< \Phi^2(\beta_{03}(t-1)) + \Big[\Phi^2(\beta_{03}t) - \Phi(\beta_{03}(t-1))\Big]^2 \end{aligned} \tag{A.5}$$

Now let $G : \mathbb{R}_+ \to \mathbb{R}_+$ be the function defined by:

$$G(t) = \frac{\Phi^2(e^*(t)) - \Phi^2(\beta_{03}(t-1))}{[\Phi(\beta_{03}t) - \Phi(\beta_{03}(t-1))]^2}$$

Combining (A.4) and (A.5), we can establish the desired result if we can show that there exists some t_0 such that $G(t) > 1$ for all $t \geq t_0$. We will actually show the stronger result $\lim_{t\to\infty} G(t) = \infty$. To keep the notation simple, we assume from here on that $\beta_{03} = 1$, the proof can be easily adapted to the general case with $\beta_{03} > 0$.

First note that:

$$\begin{aligned}\lim_{t\to\infty} G(t) &= \lim_{t\to\infty} \frac{\Phi(\hat{e}(t)) - \Phi(t-1)}{[\Phi(t) - \Phi(t-1)]^2} \cdot \Big(\Phi(\hat{e}(t)) + \Phi(t-1)\Big) \\ &= 2 \cdot \lim_{t\to\infty} \frac{\Phi(\hat{e}(t)) - \Phi(t-1)}{[\Phi(t) - \Phi(t-1)]^2}\end{aligned}$$

Both the numerator and the denominator converge to 0, hence we would like to apply L'Hôpital's rule (further ahead we show that this can indeed be done, since all the limits are well defined). For that purpose, recall that $\hat{e}(t)$ is a (symmetric) solution to (1.13) with $x = x(t)$, i.e.,

$$\hat{e}(t) = t - \Phi(\hat{e}(t)).$$

Therefore, taking implicit derivatives with respect to t, it follows that $\hat{e}$ is differentiable and:

$$\hat{e}'(t) = \frac{1}{1 + \phi(\hat{e}(t))},$$

where ϕ denotes the standard normal p.d.f.. Hence, applying L'Hôpital's rule, and using the fact that $\phi'(x) = -x \cdot \phi(x)$, it follows that:

$$\begin{aligned}\lim_{t\to\infty} G(t) &= 2 \cdot \lim_{t\to\infty} \frac{\phi(\hat{e}(t))\hat{e}'(t) - \phi(t-1)}{2[\Phi(t) - \Phi(t-1)][\phi(t) - \phi(t-1)]} \\ &= \lim_{t\to\infty} \left(\frac{\phi(t-1) - \phi(\hat{e}(t))\hat{e}'(t)}{\phi^2(t-1)}\right)\left(\frac{\phi(t-1)}{\Phi(t) - \Phi(t-1)}\right)\left(\frac{\phi(t-1)}{\phi(t-1) - \phi(t)}\right) \\ &\geq \liminf_{t\to\infty} \frac{\phi(t-1) - \phi(\hat{e}(t))\hat{e}'(t)}{\phi^2(t-1)} \times \lim_{t\to\infty} \frac{\phi(t-1)}{\Phi(t) - \Phi(t-1)} \cdots \\ &\qquad \cdots \times \lim_{t\to\infty} \frac{\phi(t-1)}{\phi(t-1) - \phi(t)}.\end{aligned}$$

We will show that the first term is strictly positive, the second limit equals $+\infty$, and the third one equals 1. This implies that $\lim_{t\to\infty} G(t) = +\infty$, thus completing the proof.

We begin with the second and third limits. First note that:

$$\frac{\phi(t)}{\phi(t-1)} = \exp\left(-\frac{1}{2}t^2\right)\exp\left(\frac{1}{2}(t-1)^2\right) = \exp\left(\frac{1}{2}-t\right) \xrightarrow[t\to\infty]{} 0.$$

This implies that:

$$\lim_{t\to\infty} \frac{\phi(t-1)}{\phi(t-1)-\phi(t)} = \lim_{t\to\infty} \frac{1}{1-\phi(t)/\phi(t-1)} = 1. \tag{A.6}$$

For the second limit, L'Hôpital's rule and (A.6) imply that:

$$\lim_{t\to\infty} \frac{\phi(t-1)}{\Phi(t)-\Phi(t-1)} = \lim_{t\to\infty} (t-1) \times \frac{\phi(t-1)}{\phi(t-1)-\phi(t)} = +\infty.$$

The first limit is slightly more complicated because we do not have a closed form expression for $\hat{e}(t)$. Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be the function defined by:

$$g(t) = \frac{\phi(t-1) - \phi(\hat{e}(t))\hat{e}'(t)}{\phi^2(t-1)} = \frac{1 - \frac{\phi(\hat{e}(t))}{\phi(t-1)}\hat{e}'(t)}{\phi(t-1)}$$

It only remains to show that $\liminf_{t\to\infty} g(t) > 0$. Suppose towards a contradiction that this is false. It is straightforward to see that $g(t) \geq 0$, and thus this is equivalent to supposing that $\liminf_{t\to\infty} g(t) = 0$.

Under this supposition, there would exist an increasing sequence (t_n) such that $\lim t_n = \infty$ and $\lim g(t_n) = 0$. Now consider the sequence (r_n) given by:

$$r_n = \frac{\phi(\hat{e}(t_n))}{\phi(t_n - 1)}$$

Since $\hat{e}(t)$ belongs to $(t-1, t)$ for all t, it follows that $r_n \in (0,1)$ for all n. Further suppose towards a contradiction that $\liminf r_n < 1$. This would imply that there exists a subsequence (t_{m_n}) such that $r_{m_n} \longrightarrow r^* \in [0,1)$. This however would imply the contradiction:

$$\lim g(t_{m_n}) = \lim \frac{1 - r_{m_n}\hat{e}'(t_{m_n})}{\phi(t_{m_n} - 1)} = +\infty > 0,$$

where we used the facts that $\lim \hat{e}'(t_{m_n}) = 1$ and $\lim \phi(t_{m_n} - 1) = 0$. Therefore, if $\lim g(t_n) = 0$, then it must be the case that $\liminf r_n \geq 1$, and thus, since $r_n \in (0,1)$, it would follow that $\lim r_n = 1$. However, this would in turn imply the contradiction:

$$\lim g(t_n) = \lim \frac{1 - r_n \hat{e}'(t_n)}{\phi(t_n - 1)} \geq \lim \frac{1 - \hat{e}'(t_n)}{\phi(t_n - 1)} = \lim \frac{1 - \frac{1}{1+\phi(\hat{e}(t_n))}}{\phi(t_n - 1)}$$
$$= \lim \frac{\phi(\hat{e}(t_n))}{\phi(t_n - 1)} \times \frac{1}{1 + \phi(\hat{e}(t_n))} = \lim r_n \times \lim \frac{1}{1 + \phi(\hat{e}(t_n))} = 1 > 0.$$

Therefore it must be the case that $\liminf_{t\to\infty} g(t) > 0$, and the proof is thus complete.

■

Appendix B

Omitted proofs from Chapter 2

B.1. Notation and definitions

$$
\begin{aligned}
[d] &= \max\{n \in \mathbb{N}, d \geq n\}, \\
\|\beta\|_e &= \sqrt{\beta^{\mathsf{T}}\beta}, \\
\|h\|_s^2 &= \max_{y \in Y} \sum_{|\lambda| \leq \kappa + \kappa_0} \int [D^\lambda h(y, x, e)]^2 dx de, \\
\|h\|_c &= \max_{y \in Y} \max_{|\lambda| \leq \kappa} \sup_{x, e} \left| D^\lambda h(y, x, e) \right|, \\
\|\theta\|_s &= \|\beta\|_e + \|h\|_s, \\
\|\theta\|_c &= \|\beta\|_e + \|h\|_c, \\
\Theta_0 &= \text{the identified set}, \\
p(\theta) &= p(\cdot|\cdot, \theta), \\
p_0 &= p(\theta_0), \quad \theta_0 \in \Theta_0, \\
P_0 &= p_0 f_x, \\
L^2(P_0) &= \left\{ g : Y \times X \to \mathbb{R} : \int g^2 dP_0 < \infty \right\}, \\
\langle g_1, g_2 \rangle_{L^2(P_0)} &= \int g_1 g_2 dP_0, \\
\|g\|_{L^2(P_0)} &= \sqrt{\langle g, g \rangle_{L^2(P_0)}}, \\
\chi^2(P_1, P_2) &= \int \left(\frac{P_1}{P_2} - 1 \right)^2 P_2 d\mu, \ \text{where } \mu \text{ is a dominating } \sigma\text{-finite positive measure},
\end{aligned}
$$

$$
\begin{aligned}
\chi^2(\theta,\theta_0) &= \chi^2(p(\theta),p_0) = \int\left(\frac{p(\theta)}{p_0}-1\right)^2 P_0 d\mu = \left\|\frac{p(\theta)}{p_0}-1\right\|^2_{L^2(P_0)},\\
H^2(P_1,P_2) &= \int(\sqrt{P_1}-\sqrt{P_2})^2 d\mu,\\
\mathcal{G} &= \left\{g\in L^2(P_0) : \|g\|_c < C\right\},\ \text{for some known } C<\infty,\\
\bar{\mathcal{G}} &= \left\{g\in\mathcal{G} : \int g dP_0 = 0, \|g\|_{L^2(P_0)} = 1\right\},\\
\mu_n\{g\} &= n^{-1}\sum_{i=1}^n g(\mathbf{y}_i,\mathbf{x}_i),\\
\text{NE} &= \{\theta\in\Theta : \theta \text{ satisfies NE constraints}\},\\
\text{NE}_j &= \Theta_0 \bigcap \{\theta\in\Theta : m_j(\theta)=0\},
\end{aligned}
$$

B.2. SOR and NE constraints

Let

$$u_i(y,x,e,\beta) = (v_i(x,\beta) - \tilde{v}_i(x,\beta)y_{-i} - e_i)y_i.$$

For $i=1,2$ denote

$$\underline{v}_i(x,b) = v_i(x,\beta) - \tilde{v}_i(x,\beta).$$

Then under the SOR assumption

$$
\begin{aligned}
A((0,0),x,\beta) =& \{e\in E \mid e_i > v_i(x,\beta), i=1,2\},\\
A((1,0),x,\beta) =& \{e\in E \mid e_1 < \underline{v}_1(x,b), e_2 > \underline{v}_2(x,b)\} \bigcup\\
& \{e\in E \mid e_1 < v_1(x,\beta), e_2 > v_2(x,\beta)\},\\
A((1,1),x,\beta) =& \{e\in E \mid e_i < \underline{v}_i(x,b), i=1,2\},\\
A^M(x,\beta) =& \{e\in E \mid \underline{v}_i(x,b) \le e_i \le v_i(x,\beta), i=1,2\}.
\end{aligned}
$$

Next, I derive a closed form solution for the Nash constraints. First, I characterize the set of parameters that are consistent with Nash behavior in terms of one equality and three inequality constraints that have to hold uniformly over the region of multiplicity. Second, I show how these infinite dimensional equality/inequality constraints can be compressed into the set of finite dimensional constraints.

In the region of multiplicity there are three NE, two in pure strategies ($(1,0)$ and $(0,1)$) and one in mixed strategies. For all β and (e,x) let $\alpha(y,x,e,\beta)$ be the probability that y is played in the mixed strategy NE in the region of multiplicity. That is,

$$\begin{aligned}\alpha((0,0),x,e,\beta) =&\frac{(e_1-\underline{v}_1(x,b))(e_2-\underline{v}_2(x,b))}{\tilde{v}_1(x,\beta)\tilde{v}_2(x,\beta)},\\ \alpha((1,1),x,e,\beta) =&\frac{(v_1(x,\beta)-e_1)(v_2(x,\beta)-e_2)}{\tilde{v}_1(x,\beta)\tilde{v}_2(x,\beta)},\\ \alpha((1,0),x,e,\beta) =&\frac{(e_1-\underline{v}_1(x,b))(v_2(x,\beta)-e_2)}{\tilde{v}_1(x,\beta)\tilde{v}_2(x,\beta)},\\ \alpha((0,1),x,e,\beta) =&\frac{(e_2-\underline{v}_2(x,b))(v_1(x,\beta)-e_1)}{\tilde{v}_1(x,\beta)\tilde{v}_2(x,\beta)}.\end{aligned}$$

Let $h(x,e)=(h(y,x,e))_{y\in Y}$ and $\alpha(x,e,\beta)=(\alpha(y,x,e,\beta))_{y\in Y}$. Then according to (2.4), for $\theta=(\beta,h)$ to be consistent with the NE assumption, the following has to be true uniformly over the multiplicity region:

$$h(x,e)\in \operatorname{co}\Big((0,0,1,0)^{\mathsf{T}},(0,0,0,1)^{\mathsf{T}},\alpha(x,e,\beta)\Big).$$

The above convex hull is easy to characterize. It is an intersection of three half-spaces and one hyperplane in $\mathbb{R}^3$. For $j=0,1,2,3$ define $\tilde{m}_j: X\times E\times\Theta\to\mathbb{R}$ such that

$$\begin{aligned}\tilde{m}_0(x,e,\theta) =&\alpha((1,1),x,e,\beta)h((0,0),x,e)-\alpha((0,0),x,e,\beta)h((1,1),x,e),\\ \tilde{m}_1(x,e,\theta) =&-\alpha((1,0),x,e,\beta)h((1,1),x,e)+\alpha((1,1),x,e,\beta)h((1,0),x,e),\\ \tilde{m}_2(x,e,\theta) =&\frac{\alpha((1,1),x,e,\beta)+\alpha((1,0),x,e,\beta)-1}{\alpha((0,0),x,e,\beta)}h((0,0),x,e)-h((1,1),x,e)-h((1,0),x,e)+1,\\ \tilde{m}_3(x,e,\theta) =&h((1,0),x,e)+h((0,1),x,e).\end{aligned}$$

Then

$$\begin{aligned}&h(x,e)\in \operatorname{co}\Big((0,0,1,0)^{\mathsf{T}},(0,0,0,1)^{\mathsf{T}},\alpha(x,e,\beta_{\mathrm{NE}})\Big)\Leftrightarrow\\ &\tilde{m}_0(x,e,\theta)=0,\quad \tilde{m}_1(x,e,\theta)\geq 0,\quad \tilde{m}_2(x,e,\theta)\geq 0\quad \tilde{m}_3(x,e,\theta)\geq 0.\end{aligned}$$

The above NE constraints have to be satisfied for almost all e and x such that $e\in A^M(x,\beta)$. The constraints are infinite-dimensional. In order to turn them into finite-dimensional constraints I follow Andrews and Shi (2013) and define the set of boxes in $\mathbb{R}^{d_e+d_x}$ with centers at $(e^\intercal, x^\intercal)^\intercal \in \mathbb{R}^{d_e+d_x}$ and side length less than $2\bar{r}$. Let $z = (e^\intercal, x^\intercal)^\intercal$. Then

$$\mathcal{C}_{\text{box}} = \Big\{ C_{z,r} = \times_{u=1}^{d_z} (z_u - r_u, z_u + r_z] \in [0,1]^{d_z} : u \le d_z, z_u \in [0,1], r_u \in (0, \bar{r}) \Big\}$$

and the set of indicator functions on this boxes

$$\mathcal{G}_{\text{box}} = \Big\{ g_{z,r} : g_{z,r}(z) = \mathbb{1} z \in C \text{ for } C \in \mathcal{C}_{\text{box}} \Big\},$$

Let $Q_{z,r}$ be the uniform distributions on $[0,1]^{d_z} \times (0, \bar{r})^{d_z}$.

Take any twice continuously differentiable with respect to β function $\kappa : E \times X \times B \to \mathbb{R}$ such that $\kappa(\cdot, \cdot, \beta)$ is strictly positive on the interior of $A^M(x, \beta)$ and equals to zero outside of it. For instance,

$$\kappa(e, x, \beta) = \times_{i=1}^{2} \kappa_1(e_i - v_i(e, x, \beta) + \tilde{v}_i(e, x, \beta)) \kappa_1(-e_i + v_i(e, x, \beta)),$$

where $\kappa_1(x) = (\max\{x, 0\})^4$. For $j = 0, 1, 2, 3$ define $m_j : \Theta \to \mathbb{R}$ such that

$$m_j(\theta) = \int_{[0,1]^{d_z} \times (0,\bar{r})^{d_z}} \mathbb{E}\big[\, g_{z,r}(\Phi(\mathbf{e}), \Phi(\mathbf{x})) \kappa(\mathbf{e}, \mathbf{x}, \beta) \tilde{m}_j(\mathbf{x}, \mathbf{e}, \theta) \,\big] \, dQ_{z,r}(z, r),$$

where $\Phi(\cdot)$ is the standard normal c.d.f..

So, I translated infinitely many equalities/inequalities to a finite set of constraints. That is,

$$\begin{aligned} & h(x, e) \in \text{co}\Big((0,0,1,0)^\intercal, (0,0,0,1)^\intercal, \alpha(x, e, \beta)\Big) \text{ a.s.} \Leftrightarrow \\ & m_0(\theta) = 0, \quad m_1(\theta) \ge 0, \quad m_2(\theta) \ge 0, \quad m_3(\theta) \ge 0. \end{aligned}$$

The above constraints have nice properties: they are smooth in β and affine in h.

B.3. Proof of Theorem 2.2

Assumptions of Theorem 3.1 in CTT are satisfied. Hence, 1 follows from Theorem 3.1 in CTT.

For the conclusion 2, note that Theorem 3.2 in CTT states that the rate of convergence of the sieve MLE in terms of the Pearson distance is determined by the sieve approximation error under the squared Pearson distance and the measure of sieve model complexity in terms of the Hellinger distance with bracketing.

Define $\Gamma : B \times W^s(Y \times X \times E') \to L^2(P_0)$ as

$$\Gamma(\beta, v)(y, x) = \int_{A^M(\beta,x)} v(y, x, e) f_{e|x}(e|x, \beta) de.$$

Note that p_0 is bounded away from zero by compactness of X. Hence, due to the mean value theorem there exists a constant $0 < C_1 < \infty$ such that the sieve approximation error under the squared Pearson distance satisfies the following inequality:

$$\begin{aligned}\inf_{\theta \in \Theta_{k(n)}} \left\| \frac{p(\theta)}{p_0} - 1 \right\|^2_{L^2(P_0)} &\leq \inf_{h \in \mathcal{H}_{k(n)-d_\beta}} \left\| \frac{\Gamma(\beta_0, h - h_0)}{p_0} \right\|^2_{L^2(P_0)} \\ &\leq C_1 \inf_{h \in \mathcal{H}_{k(n)-d_\beta}} \|h - h_0\|^2_\infty = O(\gamma_n^2),\end{aligned}$$

where the last equality follows from (2.7). As a result, the sieve approximation error under the squared Pearson distance is $O(\gamma_n)$.

Next, if I show that the measure of sieve model complexity in terms of the Hellinger distance with bracketing is of the order $\sqrt{k(n)/n}$, then the 2 follows from Theorem 3.2 of CTT. Recall that $\Theta_{k(n)}$ is of finite dimension and the probability density is Lipschitz in $\theta \in \Theta_{k(n)}$. Moreover, because of restrictions on derivatives of h and B, the parameter space (β and coefficients in front of basis functions) is a bounded subset of $\mathbb{R}^{k(n)}$. Hence, by Theorem 2.7.11 in Van Der Vaart and Wellner (1996), there exists a constant $0 < C < \infty$ such that

$$N_{[]}\Big(u, \{p(\theta) \mid \ \theta \in \Theta_{k(n)}\}, H(\cdot, \cdot)\Big) \leq \left(\frac{C}{u}\right)^{k(n)},$$

where $N_{[]}\Big(u, \mathcal{F}_n, \|\cdot\|\Big)$ is the bracketing number for the set $\mathcal{F}_n$ with respect to the norm $\|\cdot\|$.

As a result, the bound for the sieve measure of complexity in terms of the Hellinger distance with bracketing, ξ_n, should satisfy

$$\xi_n^{-2} \int_{2^{-8}\xi_n^2}^{\sqrt{2}\xi_n} \sqrt{\log\left(\frac{C}{u}\right)^{k(n)}} du \leq \sqrt{k(n)}\xi_n^{-1} = O\left(\sqrt{n}\right),$$

which is ensured by $\xi_n = O\left(\sqrt{k(n)/n}\right)$.

Combining the results for the sieve approximation error under the squared Pearson distance and the measure of sieve model complexity in terms of the Hellinger distance with bracketing I get that $\chi(\hat{\theta}_n, \theta_0) = O_p(\delta_n)$ for all $\hat{\theta}_n \in \hat{\Theta}_n$ and $\theta_0 \in \Theta_0$, where $\delta_n = \max\left\{(k(n)/n)^{1/2}, \gamma_n\right\}$.

Note that by the consistency result there exists $\epsilon_n = o(1)$, such that

$$\inf_{\theta_0 \in \Theta_0} \left\|\hat{\theta}_n - \theta_0\right\|_c = o_p(\epsilon_n),$$
$$\|\theta - \theta_0\|_c \leq \epsilon_n \Rightarrow \chi(\theta, \theta_0) = O(\delta_n \log\log(n))$$

Let $B_n(\theta_0) = \{\theta \in \Theta_{k(n)} : \|\theta - \theta_0\|_c < \epsilon_n, \forall \theta_0', \|\theta - \theta_0\|_c \leq \|\theta - \theta_0'\|_c\}$ and θ_{0n} be a projection of θ_0 on $\Theta_{k(n)}$. That is, I am considering points in a shrinking neighborhood of θ_0 such that θ_0 is the closest point in the identified set. The following lemma establishes the rate of convergence of the constrained sieve MLE to the closest point in $\Theta_0 \bigcap \mathrm{NE}$.

Lemma B.1 *Under Assumptions 1, 2, 3 and 4, if* $\delta_n \log\log(n) = o(n^{-1/4})$ *then uniformly in* $\theta_0 \in \Theta_0 \bigcap \mathrm{NE}$,

$$\sup_{\theta \in B_n(\theta_0)} \|\beta - \beta_0\|_{\theta_0} = o_p(n^{-1/4}),$$
$$\sup_{\theta \in B_n(\theta_0)} \left\|\frac{\Gamma(\beta_0, h - h_0)}{p_0}\right\|_{L^2(P_0)} = o_p(n^{-1/4}).$$

Proof. Let θ_{0n} be a projection of θ_0 on $\Theta_{k(n)}$. By Theorem 2.2.1-2 and the assumption that $\delta_n \log\log(n) = o(n^{-1/4})$

$$\|h_{0n} - h_0\|_\infty = O(\gamma_n) = O(\delta_n) = o(n^{-1/4}).$$

Now, by Jensen's inequality, there exists a finite constant $C_1 > 0$ such that

$$\left\|\frac{\Gamma(\beta, h_{0n} - h_0)}{p_0}\right\|_{L^2(P_0)} \leq C_1 \|h_{0n} - h_0\|_\infty = o(n^{-1/4}).$$

Then, since each element of the matrix $\partial_{\beta\beta^\intercal}(p(y|x,\beta))/p_0(y|x)$ is uniformly bounded, and by the mean value theorem the following holds

$$o(n^{-1/4}) = \chi((\beta, h_0), \theta_0) = \left\|\frac{p(\beta, h_0)}{p_0} - 1\right\|_{L^2(P_0)} =$$

$$= \left\| \frac{\partial_{\beta^\intercal} p(\theta_0)}{p_0}(\beta - \beta_0) + O\left(\|\beta - \beta_0\|_e^2\right) \right\|_{L^2(P_0)} \geq$$

$$\geq |\text{the triangular inequality}| \geq \left| \left\| \frac{\partial_{\beta^\intercal} p(\theta_0)}{p_0}(\beta - \beta_0) \right\|_{L^2(P_0)} - O\left(\|\beta - \beta_0\|_e^2\right) \right| =$$

$$= \left| \|\beta - \beta_0\|_{\theta_0} - O\left(\|\beta - \beta_0\|_{\theta_0}^2\right) \right|.$$

Hence, $\sup_{\theta \in B_n(\theta_0)} \|\beta - \beta_0\|_{\theta_0} = o(n^{-1/4})$.

Similarly, since $\left\| \partial_{\beta^\intercal} \Gamma(\beta_0, h - h_0) \right\|_{L^2(P_0)} = o(1)$, by the mean value theorem and the triangular inequality

$$o(n^{-1/4}) = \chi(\theta, \theta_0) =$$

$$= \left\| \frac{\Gamma(\beta_0, h - h_0)}{p_0} + \frac{\partial_{\beta^\intercal} \Gamma(\beta_0, h - h_0)(\beta - \beta_0)}{p_0} + \frac{\partial_{\beta^\intercal} p(\theta_0)}{p_0}(\beta - \beta_0) + o_p(n^{-1/2}) \right\|_{L^2(P_0)} \geq$$

$$\geq \left| \left\| \frac{\Gamma(\beta_0, h - h_0)}{p_0} \right\|_{L^2(P_0)} - o(n^{-1/4}) \right|.$$

As a result,

$$\sup_{\theta \in B_n(\theta_0)} \left\| \frac{\Gamma(\beta_0, h - h_0)}{p_0} \right\|_{L^2(P_0)} = o(n^{-1/4}).$$

■

B.4. Auxiliary results

In this appendix I state and prove the auxiliary lemmas used to prove the main results. Since, $p_0(y|x)$ is the conditional density of $\mathbf{y}|\mathbf{x}$, the underlying measure dP_0 is equal to $p_0 f_x dx$, where $f_x(x)$ is a p.d.f. of $\mathbf{x}$.

Consider $\mathcal{G} \subseteq L^2(P_0)$ - the set of all functions that are at least $[d_x/2] + 1$-times continuously differentiable in x. Note that since X is a convex compact subset of $\mathbb{R}^{d_x}$ and Y is finite, by Corollary 2.7.2 together with the bracketing central limit theorem (Theorem 2.5.6), Theorem 2.10.1 and Theorem 2.10.6 in Van Der Vaart and Wellner

(1996), $\mathcal{G}$ is Donsker. Similarly the set of functions

$$\bar{\mathcal{G}} = \left\{ g \in \mathcal{G} : \int g dP_0 = 0, \|g\|_{L^2(P_0)} = 1 \right\}$$

is Donsker as well.

For every $\theta_0 \in \Theta_0 \bigcap \mathrm{NE}$ define $V_n(\theta_0) = span(\Theta - \{\theta_{0n}\})$, where $span(A)$ is closed linear span of A. Let $M_{n,\theta_0} : V(\theta_0) \to \mathcal{G} \subseteq L^2(P_0)$ be a linear operator such that

$$M_{n,\theta_0}[v](y,x) = \frac{\partial_\theta p(y|x,\theta_{0n})[v]}{p_0(y|x)},$$

where $v = (v_\beta, v_h)$ and $\partial_\theta p(\theta_{0n})[\cdot] = \partial_t p(\theta_{0n} + t\cdot)|_{t=0}$.

Lemma B.2 *Under assumption of Theorem 2.2*

$$\sup_{\theta_0 \in \Theta_0 \cap \mathrm{NE}} \sup_{\theta \in B_n(\theta_{0n})} \left\| \frac{p(\theta)}{p_0} - 1 - M_{n,\theta_0}[\theta - \theta_{0n}] \right\|_{L^2(P_0)} = o_p(n^{-1/2}),$$

Proof. By definition of $B_n(\theta_0)$, Lemma B.1 and linearity of $p(\theta)$ in h

$$\begin{aligned}
\frac{p(\theta)}{p_0} &= \frac{p(\beta_0, h)}{p_0} + \frac{\partial_{\beta^\intercal} p(\beta_0, h)(\beta - \beta_0)}{p_0} + o_p(n^{-1/2}) = \\
&= \frac{p(\theta_{0n})}{p_0} + \frac{\partial_\theta p(\theta_{0n})[\theta - \theta_{0n}]}{p_0} + \frac{\partial_{\beta^\intercal} \Gamma(\beta_0, h - h_{0n})(\beta - \beta_{0n})}{p_0(y|x)} + o_p(n^{-1/2}) = \\
&= 1 + M_{n,\theta_0}[\theta - \theta_{0n}] + O_p(n^{-1/4}) o_p(n^{-1/4}) + o_p(n^{-1/2}) = 1 + M_{n,\theta_0}[\theta - \theta_{0n}] + o_p(n^{-1/2}).
\end{aligned}$$

■

Recall that θ_{0n} is a projection of θ_0 on $\Theta_{k(n)}$.

Lemma B.3 *If $\gamma_n = o(n^{-1/2})$, then for every $\theta_0 \in \Theta_0 \bigcap \mathrm{NE}_j$, $j = 0,1,2,3$, there exists $g_n(\theta_{0n}) \in \bar{\mathcal{G}}$ such that*

$$\sup_{\theta \in B_n(\theta_0) \bigcap \mathrm{NE}_j} \left| \left\langle \frac{p(\theta)}{p_0} - 1, g_n(\theta_{0n}) \right\rangle_{L^2(P_0)} \right| = o_p(n^{-1/2}).$$

Proof. Fix some $j = 0,1,2,3$, $\theta_0 \in \Theta_0 \bigcap \mathrm{NE}_j$ and $\theta \in B_n(\theta_0) \bigcap \mathrm{NE}_j$. Define $B_n^j(\theta_0) = B_n(\theta_0) \bigcap \mathrm{NE}_j$. By Lemma B.1, since $m_j(\theta) = m_j(\theta_0) = 0$, and $m_j(\cdot, h)$ is twice continuously differentiable, I have the expansion

$$\begin{aligned}0 =& m_j(\theta) = m_j(\beta_0, h) + \partial_{\beta^\intercal} m_j(\beta_0, h)(\beta - \beta_0) + o_p(n^{-1/2}) \\ =& m_j(\beta_0, h) - m_j(\beta_0, h_0) + \partial_{\beta^\intercal} m_j(\beta_0, h)(\beta - \beta_0) + o_p(n^{-1/2}).\end{aligned}$$

Since $\|h_{0n} - h_0\|_\infty = O(\gamma_n) = o(n^{-1/2})$ by assumption of the lemma,

$$|m_j(\beta_0, h_0) - m_j(\beta_0, h_{0n})| \leq \|h_{0n} - h_0\|_\infty = o_p(n^{-1/2}).$$

Hence,

$$\begin{aligned}0 =& m_j(\theta) = m_j(\beta_0, h) - m_j(\beta_0, h_{0n}) + \partial_{\beta^\intercal} m_j(\beta_0, h_{0n})(\beta - \beta_0) + \\ &+ \left(\partial_{\beta^\intercal} m_j(\beta_0, h) - \partial_{\beta^\intercal} m_j(\beta_0, h_{0n})\right)(\beta - \beta_0) + o_p(n^{-1/2}).\end{aligned} \tag{B.1}$$

Note that

$$m_j(\beta_0, h) - m_j(\beta_0, h_{0n}) + \partial_{\beta^\intercal} m_j(\beta_0, h_{0n})(\beta - \beta_0) = \partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}],$$

where $\partial_\theta m_j(\theta_{0n})[\cdot] = \partial_t m_j(\theta_{0n} + t\cdot)|_{t=0}$.

Next, since for every β, $m_j(\beta, h)$ is affine in h, there exists a linear operator $\psi_j(\beta)$, such that

$$m_j(\beta_0, h) - m_j(\beta_0, h_{0n}) = \psi_j(\beta_0)[h - h_{0n}].$$

If $m_j(\beta_0, \cdot)$ is linear, then $\psi_j(\beta_0)[h - h_{0n}] = m_j(\beta_0, h - h_{0n})$.

Hence, equation (B.1) can be rewritten as

$$0 = m_j(\theta) = \partial_\theta m_j(\theta_0)[\theta - \theta_{0n}] + \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}](\beta - \beta_0) + o_p(n^{-1/2}).$$

So,

$$\sup_{\theta \in B_n^j(\theta_0)} \left|\partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}] + \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}](\beta - \beta_0)\right| = o_p(n^{-1/2}). \tag{B.2}$$

Note that by the triangular inequality,

$$\begin{aligned}&\sup_{\theta \in B_n^j(\theta_0)} \left|\partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}] + \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}](\beta - \beta_0)\right| \geq \\ &\sup_{\theta \in B_n^j(\theta_0)} \left| |\partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}]| - \left|\partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}](\beta - \beta_0)\right| \right| \geq\end{aligned}$$

$$\left| \sup_{\theta \in B_n^j(\theta_0)} |\partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}]| - \sup_{\theta \in B_n^j(\theta_0)} \left| \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}](\beta - \beta_0) \right| \right| =$$

$$\sup_{\theta \in B_n^j(\theta_0)} |\partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}]| \, |1 - \xi_n| \,, \tag{B.3}$$

where

$$\xi_n = \frac{\sup_{\theta \in B_n^j(\theta_0)} \left| \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}](\beta - \beta_0) \right|}{\sup_{\theta \in B_n^j(\theta_0)} |\partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}]|}.$$

Since by the Cauchy-Schwarz and the triangular inequalities,

$$\frac{\sup_{\theta \in B_n^j(\theta_0)} \left| \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}](\beta - \beta_0) \right|}{\sup_{\theta \in B_n^j(\theta_0)} |\partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}]|} \leq$$

$$\frac{\sup_{\theta \in B_n^j(\theta_0)} \left\| \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}] \right\|_e \sup_{\theta \in B_n^j(\theta_0)} \|\beta - \beta_0\|_e}{\sup_{\theta \in B_n^j(\theta_0)} |\partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}]|} \leq$$

$$\frac{\sup_{\theta \in B_n^j(\theta_0)} \left\| \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}] \right\|_e}{\left| \|\partial_\beta m_j(\theta_{0n})\|_e - \dfrac{\sup_{\theta \in B_n^j(\theta_0)} |\psi_j(\beta_0)[h - h_{0n}]|}{\sup_{\theta \in B_n^j(\theta_0)} \|\beta - \beta_0\|_e} \right|},$$

and

$$\begin{cases} \dfrac{\sup_{\theta \in B_n^j(\theta_0)} |\psi_j(\beta_0)[h - h_{0n}]|}{\sup_{\theta \in B_n^j(\theta_0)} \|\beta - \beta_0\|_e} \to_{n \to \infty} \infty, \\ \sup_{\theta \in B_n^j(\theta_0)} \left\| \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}] \right\|_e = o_p(1) \end{cases}$$

it follows that $\xi_n = o_p(1)$.

The fact that $\xi_n = o_p(1)$ together with (B.2) and (B.3) imply that

$$\sup_{\theta \in B_n(\theta_0) \cap \mathrm{NE}_j} |\partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}]| = o_p(n^{-1/2}). \tag{B.4}$$

Let $V_n^{null}(\theta_0) \subseteq V_n(\theta_0)$ be a kernel of the linear operator $M_{n,\theta_0}[\cdot]$ that maps to $\mathcal{G}$. Indeed, since I assume that h, h_{0n} and h_0 are at least $[d_x/2] + 1$-times continuously differentiable and $Y \times X$ is compact, $M_{n,\theta_0}[v] \in \mathcal{G}$. Let $V_n^{\perp}(\theta_0)$ be an orthogonal complement of $V_n(\theta_0)$. Note that for a sufficiently large n, $V_n^{\perp}(\theta_0) \setminus \{0\} \neq \emptyset$. Then $\langle \cdot, \cdot \rangle_{n,\theta_0}$ such that

$$\langle v_1, v_2 \rangle_{n,\theta_0} = \langle M_{n,\theta_0}[v_1], M_{n,\theta_0}[v_2] \rangle_{L^2(P_0)}$$

is a proper inner product in $V_n^{\perp}(\theta_{0n})$.

So, $\theta - \theta_{0n}$ can be decomposed into v_1 and v_2, such that $v_1 \in V^{\perp}(\theta_{0n})$ and $v_2 \in V_n^{null}(\theta_0)$. Because of the fact that $\|\theta - \theta_{0n}\|_c = o_p(1)$, equation (B.4) and the triangular inequality

$$o_p(n^{-1/2}) = \left|\partial_\theta m_j(\theta_{0n})[v_1] + \partial_{\beta^\intercal} m_j(\beta_0, v_{1,h})(\beta - \beta_0)\right|.$$

By the Riesz representation theorem for linear operators in finite dimensional spaces, there exist $v_n^*(\theta_{0n})$ different from zero, such that

$$\partial_\theta m_j(\theta_{0n})[v_1] = \langle v_n^*(\theta_{0n}), v_1\rangle_{n,\theta_0}.$$

Define $\tilde{g}_n(\theta_{0n}, \beta) = M_{n,\theta_0}[v_n^*(\theta_{0n})]$. By construction $\tilde{g}_n(\theta_{0n}) \in \mathcal{G}$. Hence,

$$g_n(\theta_{0n}) = \frac{\tilde{g}_n(\theta_{0n}) - \int \tilde{g}_n(\theta_{0n})dP_0}{\|\tilde{g}_n(\theta_{0n}) - \int \tilde{g}_n(\theta_{0n})dP_0\|_{L^2(P_0)}} \in \bar{\mathcal{G}}.$$

Note that for every constant C,

$$\left\langle \frac{p(\theta)}{p_0} - 1, C\right\rangle_{L^2(P_0)} = C\left(\int\left(\frac{p(\theta)}{p_0} - 1\right)p_0 d\mu\right) = C(1-1) = 0.$$

Then, by the Taylor expansion, Lemma B.1, the Cauchy-Schwarz inequality and the triangular inequality, for every $\theta \in B_n(\theta_0)\bigcap \mathrm{NE}_j$

$$\begin{aligned}\left|\left\langle \frac{p(\theta)}{p_0} - 1, g_n(\theta_{0n})\right\rangle_{L^2(P_0)}\right| &= \left|\left\langle \frac{p(\theta)}{p_0} - 1, \frac{\tilde{g}_n(\theta_{0n})}{\|\tilde{g}_n(\theta_{0n}) - \int \tilde{g}_n(\theta_{0n})dP_0\|_{L^2(P_0)}}\right\rangle_{L^2(P_0)}\right| = \\ &= \frac{\left|\langle M_{n,\theta_0}[\theta - \theta_{0n}], \tilde{g}_n(\theta_{0n})\rangle_{L^2(P_0)}\right|}{\|\tilde{g}_n(\theta_{0n}) - \int \tilde{g}_n(\theta_{0n})dP_0\|_{L^2(P_0)}} + o_p(n^{-1/2}) = \\ &= \frac{|\partial_\theta m_j(\theta_{0n})[v_1]|}{\|\tilde{g}_n(\theta_{0n}) - \int \tilde{g}_n(\theta_{0n})dP_0\|_{L^2(P_0)}} + o_p(n^{-1/2}) = o_p(n^{-1/2}).\end{aligned}$$

■

Lemma B.4 *Under Assumptions 1, 2 and 3, if* $\delta_n \log\log(n) = o(1)$*, then*

$$\sup_{\theta_0\in\Theta_0}\sup_{\substack{\theta\in\mathcal{B}_n(\theta_0),\\ g\in\bar{\mathcal{G}}}}\left\|\sqrt{n}\mu_n\left\{\left(\frac{p(\theta)}{p_0} - 1\right)g\right\}\right\|_e = o_p(1),$$

$$\sup_{g\in\bar{\mathcal{G}}} \|\mu_n\{g\}\|_e = O_p(n^{-1/2}).$$

Proof. Note that since $g \in \mathcal{G}$ and assumptions of the lemma

$$\left\{\left(\frac{p(\theta)}{p_0} - 1\right) g : \theta \in \mathcal{B}_n(\theta_0), g \in \bar{\mathcal{G}}\right\} \subseteq \mathcal{G}$$

By the Cauchy-Schwarz inequality

$$\left\|\left(\frac{p(\theta)}{p_0} - 1\right) g\right\|^2_{L^2(P_0)} \leq \left\|\left(\frac{p(\theta)}{p_0} - 1\right)\right\|^2_{L^2(P_0)} = \chi^2(\theta, \theta_0)$$

Hence,

$$\sup_{\substack{\theta\in\mathcal{B}_n(\theta_0),\\ g\in\bar{\mathcal{G}}}} \left\|\sqrt{n}\mu_n\left\{\left(\frac{p(\theta)}{p_0} - 1\right) g\right\}\right\|_e \leq \sup_{\substack{g\in\mathcal{G},\\ \|g\|\leq\delta_n \log\log(n)}} \left\|\sqrt{n}\mu_n\{g\}\right\|_e = o_p(1),$$

where the last equality follows from the facts that $\mathcal{G}$ is Donsker and

$$\chi(\theta, \theta_0) = O(\delta_n \log\log(n)) = o(1).$$

■

Consider a perturbation in the probability density sieve space: for every $\theta_0 \in \Theta_0$, $\theta \in \mathcal{B}_n(\theta_0)$, $g \in \bar{\mathcal{G}}$.

$$p(\theta(t_n, g)) = p(\theta) + t_n g p_0,$$

where $t_n \in \mathcal{T}_n = \{t \in [-1, 1] : \|t\|_e \leq Cn^{-1/2}\}$ for some $C < \infty$.

Let

$$R(y, x, \theta, \theta_0) = \log p(y|x; \theta) - \log p(y|x, \theta_0) - \left[\frac{p(y|x;\theta)}{p_0(y|x)} - 1\right].$$

Lemma B.5 *Under Assumptions 1, 2 and 3, if $\delta_n \log\log(n) = o(1)$, then*

$$\sup_{\theta_0\in\Theta_0} \sup_{\substack{\theta\in\mathcal{B}_n(\theta_0),\\ t\in\mathcal{T}_n,\\ g\in\bar{\mathcal{G}}}} \|\mu_n\{R(\theta, \theta_0) - R(\theta(t, g), \theta_0)\}\|_e = o_p(n^{-1})$$

Proof. Notice that for a sufficiently large n, $p(\theta)$ is close to p_0. Moreover, since $p(\theta)$

and p_0 are bounded away from zero, the following expansion holds uniformly in (y, x) and θ.

$$\log p(y|x;\theta) - \log p_0(y|x) = \log\left\{1 + \left(\frac{p(y|x;\theta)}{p_0(y|x)} - 1\right)\right\}$$
$$= \left(\frac{p(y|x;\theta)}{p_0(y|x)} - 1\right) - 1/2\left(\frac{p(y|x;\theta)}{p_0(y|x)} - 1\right)^2 + \sum_{l=1}^{\infty}\frac{(-1)^{l+1}}{l+2}\left(\frac{p(y|x;\theta)}{p_0(y|x)} - 1\right)^{l+2}$$

As a result, uniformly in θ, g and t_n

$$R(y,x,\theta,\theta_0) - R(y,x,\theta(t_n,g),\theta_0) = \log p(y|x;\theta) - \log p(y|x,\theta(t_n,g)) + t_n g(y,x) =$$
$$\left(\frac{p(y|x;\theta)}{p_0(y|x)} - 1\right) t_n g(y,x) + (t_n g(y,x))^2/2$$
$$+ \sum_{l=1}^{\infty}\frac{(-1)^{l+1}}{l+2}\left[\left(\frac{p(y|x;\theta)}{p_0(y|x)} - 1\right)^{l+2} - \left(\frac{p(y|x,\theta(t_n,g))}{p_0(y|x)} - 1\right)^{l+2}\right],$$

and $\left(\frac{p(y|x;\theta)}{p_0(y|x)} - 1\right) tg(y,x)$ is the leading term.

Note that since $g \in \bar{\mathcal{G}}$ and $t_n \leq Cn^{-1/2}$ there exists a constant $0 < C_1 < \infty$ such that

$$\sup_{\substack{\theta\in\mathcal{B}_n(\theta_0),\\ t_n\in\mathcal{T}_n,\\ g\in\bar{\mathcal{G}}}} \left| n\mu_n\left\{\left(\frac{p(\theta)}{p_0} - 1\right) t_n g\right\}\right| \leq$$
$$\sup_{\substack{\theta\in\mathcal{B}_n(\theta_0),\\ g\in\bar{\mathcal{G}}}} C_1\left\|\sqrt{n}\mu_n\left\{\left(\frac{p(\theta)}{p_0} - 1\right) g\right\}\right\|_e = o_p(1),$$

where the last equality follows from Lemma B.4. ■

Lemma B.6 *Under Assumptions 1, 2 and 3, if $\delta_n \log\log(n) = o(1)$, then uniformly in $\theta_0 \in \Theta_0$, $\theta \in \mathcal{B}_n(\theta_0)$, $g \in \bar{\mathcal{G}}$ and $t_n \in \mathcal{T}_n$*

$$n^{-1}\left(L_n(\theta(t_n,g)) - L_n(\theta)\right) =$$
$$t_n\mu_n\{g\} - \frac{\|t_n\|_e^2}{2} - t_n\left\langle\frac{p(\theta)}{p_0} - 1, g\right\rangle_{L^2(P_0)}(1 + o(1)) + o_p(n^{-1}).$$

Proof. Recall that

$$R(y,x,\theta,\theta_0) = \log p(y|x;\theta) - \log p(y|x,\theta_0) - \left[\frac{p(y|x;\theta)}{p_0(y|x)} - 1\right].$$

Then, since $\mathbb{E}[\log p(\mathbf{y}|\mathbf{x};\theta) - \log p_0(\mathbf{y}|\mathbf{x})] = -K(p(\theta), p_0) = -\frac{1}{2}\chi^2(\theta,\theta_0)(1+o(1))$ (see Remark 3.2 in CTT),

$$\frac{1}{n}\sum_{i=1}^{n}\{\log p(\mathbf{y}_i|\mathbf{x}_i,\theta) - \log p_0(\mathbf{y}_i|\mathbf{x}_i)\} = \mu_n\{\log p(\theta) - \log p_0\} + \mathbb{E}[\log p(\mathbf{y}|\mathbf{x};\theta) - \log p_0(\mathbf{y}|\mathbf{x})] =$$
$$\mu_n\{R(\theta,\theta_0)\} + \mu_n\left\{\frac{p(\theta)}{p_0} - 1\right\} - \frac{1}{2}\chi^2(\theta,\theta_0)(1+o(1)).$$

Hence,

$$\frac{L_n(\theta(t_n,g)) - L_n(\theta)}{n} =$$
$$\frac{1}{n}\sum_{i=1}^{n}\{\log p(\mathbf{y}_i|\mathbf{x}_i,\theta(t_n,g)) - \log p_0(\mathbf{y}_i|\mathbf{x}_i)\} - \frac{1}{n}\sum_{i=1}^{n}\{\log p(\mathbf{y}_i|\mathbf{x}_i,\theta) - \log p_0(\mathbf{y}_i|\mathbf{x}_i)\} =$$
$$-\frac{1}{2}\left\{\chi^2(\theta(t_n,g),\theta_0) - \chi^2(\theta,\theta_0)\right\}(1+o(1)) + \mu_n\left\{\frac{p(\theta(t_n,g)) - p(\theta)}{p_0}\right\} +$$
$$\mu_n\{R(\theta(t_n,g),\theta_0) - R(\theta,\theta_0)\} = (1) + (2) + (3).$$

Next, note that

$$\begin{aligned}(1) =& -\frac{1}{2}\left\{\chi^2(\theta(t_n,g),\theta_0) - \chi^2(\theta,\theta_0)\right\}(1+o(1)) = \\ & -\frac{1}{2}\left\{\left\|\frac{p(\theta(t_n,g)}{p_0} - 1\right\|^2_{L^2(P_0)} - \left\|\frac{p(\theta)}{p_0} - 1\right\|^2_{L^2(P_0)}\right\}(1+o(1)) = \\ & -\frac{\|t_n\|_e^2}{2} - t_n\left\langle\frac{p(\theta)}{p_0} - 1, g\right\rangle_{L^2(P_0)}(1+o(1)) + o(n^{-1}),\end{aligned}$$

$$(2) = \mu_n\left\{\frac{p(\theta(t_n,g)) - p(\theta)}{p_0}\right\} = t_n\mu_n\{g\},$$

$$(3) = \mu_n\{R(\theta(t_n,g),\theta_0) - R(\theta,\theta_0)\} = o_p(n^{-1}) \text{ (see Lemma B.5)}.$$

■

Recall that $\hat{\Theta}_n$ is the set of unconstrained sieve MLEs.

Lemma B.7 *Under Assumptions 1, 2 and 3*

$$\sup_{\hat{\theta}_n \in \hat{\Theta}_n} \sup_{g \in \bar{\mathcal{G}}} \left\| \left\langle \frac{p(\hat{\theta}_n)}{p_0} - 1, g \right\rangle_{L^2(P_0)} (1 + o(1)) - \mu_n \{g\} \right\|_e = o_p(n^{-1/2}).$$

Proof. Take any positive sequence t_n^* such that $\|t_n^*\|_e = o(n^{-1/2})$ and $\hat{\theta}_n(t_n^*, g) \in \mathcal{B}_n(\theta_0)$. By the definition of $\hat{\theta}_n$ and Lemma B.6

$$- o_p(n^{-1}) \leq n^{-1}\{L_n(\hat{\theta}_n) - L_n(\hat{\theta}_n(t_n^*, g))\} =$$
$$t_n^* \left[\left\langle \frac{p(\hat{\theta}_n)}{p_0} - 1, g \right\rangle_{L^2(P_0)} (1 + o(1)) - \mu_n \{g\} \right] + o_p(n^{-1})$$

Similarly, if one takes $-t_n^*$, then

$$- o_p(n^{-1}) \leq n^{-1}\{L_n(\hat{\theta}_n) - L_n(\hat{\theta}_n(-t_n^*, g))\} =$$
$$- t_n^* \left[\left\langle \frac{p(\hat{\theta}_n)}{p_0} - 1, g \right\rangle_{L^2(P_0)} (1 + o(1)) - \mu_n \{g\} \right] + o_p(n^{-1})$$

Hence,

$$- o_p(n^{-1}) \leq \pm t_n^* \left[\left\langle \frac{p(\hat{\theta}_n)}{p_0} - 1, g \right\rangle_{L^2(P_0)} (1 + o(1)) - \mu_n \{g\} \right] + o_p(n^{-1})$$

The later and the fact that $\|t_n^*\|_e = o(n^{-1/2})$ imply the statement of the lemma. ■

B.5. Proof of Theorem 2.3

Without loss of generality I consider the case when only the equality constraint is binding. The proof consists of three steps.

Step1. Recall that $\hat{\Theta}_n$ and $\tilde{\Theta}_n$ are the sets of unconstrained and constrained sieve MLEs respectively. Take an arbitrary $\theta_0 \in \Theta_0 \bigcap \mathrm{NE}$. Consider a perturbation in the probability density sieve space: for $g_n(\theta_{0n})$ from Lemma B.3

$$p(\theta(t_n, g_n(\theta_{0n}))) = p(\theta) + t_n g_n(\theta_{0n}) p_0.$$

By Lemmas B.6 and B.3

$$n^{-1}[L_n(\theta_n(t_n, g_n(\theta_{0n}))) - L_n(\theta_n)] = t_n\mu_n\left\{g_n(\theta_{0n})\right\} - \frac{t_n^2}{2} + o_p(n^{-1})$$

Then by Lemma B.4 and properties of the quadratic forms, the maximizer belongs to $\mathcal{T}_n$ and is equal to $\mu_n\left\{g_n(\theta_{0n})\right\}$. Hence,

$$\sup_{t_n \in \mathcal{T}_n} n^{-1}[L_n(\theta(t_n, g_n(\theta_{0n}))) - L_n(\theta)] - o_p(n^{-1}) = \frac{\left\|\mu_n\left\{g_n(\theta_{0n})\right\}\right\|_e^2}{2} + o_p(n^{-1})$$

As a result,

$$\begin{aligned}
&n^{-1}[L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n)] \geq \sup_{\theta_0 \in \Theta_0 \bigcap \mathrm{NE}} \inf_{\theta \in B_n(\theta_0)} n^{-1}[L_n(\hat{\theta}_n) - L_n(\theta)] + o_p(n^{-1}) \geq \\
&\sup_{\theta_0 \in \Theta_0 \bigcap \mathrm{NE}} \inf_{\theta \in B_n(\theta_0)} \sup_{t_n \in \mathcal{T}_n} n^{-1}[L_n(\theta(t_n, g_n(\theta_{0n}))) - L_n(\theta)] + o_p(n^{-1}) = \\
&\sup_{\theta_0 \in \Theta_0 \bigcap \mathrm{NE}} \frac{\left\|\mu_n\left\{g_n(\theta_{0n})\right\}\right\|_e^2}{2} + o_p(n^{-1}).
\end{aligned}$$

Finally,

$$n^{-1}[L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n)] \geq \sup_{\theta_0 \in \Theta_0 \bigcap \mathrm{NE}} \frac{\left\|\mu_n\left\{g_n(\theta_{0n})\right\}\right\|_e^2}{2} + o_p(n^{-1}) \qquad \text{(B.5)}$$

Step 2. Consider the following perturbation of $\hat{\theta}_n$: t_n^* and $g_n(\theta_{0n})$ such that

$$\begin{aligned}
t_n^* &= -\left\langle \frac{p(\hat{\theta}_n)}{p_0} - 1, g_n(\theta_{0n}) \right\rangle_{L^2(P_0)} + \epsilon_n, \\
p(\theta^*) &= p(\hat{\theta}_n) + t_n^* g_n(\theta_0) p_0,
\end{aligned}$$

where $\epsilon_n = o_p(n^{-1/2})$. Assume for a moment that for some ϵ_n, $\theta^* \in B_n(\theta_0) \bigcap \mathrm{NE}$. Then, since $t_n^* = O_p(n^{-1/2})$ (Lemma B.7 and Lemma B.4) and by applying Lemma B.6 and Lemma B.7 to t^* and $g_n(\theta_{0n})$, one can get that

$$n^{-1}[L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n)] \leq n^{-1}[L_n(\hat{\theta}_n) - L_n(\theta^*)] \leq \frac{\|t_n^*\|_e^2}{2} + o_p(n^{-1})$$

After applying Lemma B.7 to t^* and $g_n(\theta_0)$, I get that

$$n^{-1}[L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n)] \leq \sup_{\theta_0 \in \Theta_0 \bigcap \mathrm{NE}} \frac{\|\mu_n\{g_n(\theta_{0n})\}\|_e^2}{2} + o_p(n^{-1}). \qquad \text{(B.6)}$$

It remains to show that such ϵ_n exists. Note that

$$\left\langle \frac{p(\theta^*)}{p_0} - 1, g_n(\theta_0) \right\rangle_{L^2(P_0)} = -\left\langle \frac{p(\hat{\theta}_n)}{p_0} - 1, g_n(\theta_0) \right\rangle_{L^2(P_0)} + \left\langle \frac{p(\hat{\theta}_n)}{p_0} - 1, g_n(\theta_0) \right\rangle_{L^2(P_0)} + \epsilon_n = \epsilon_n = o_p(n^{-1}$$

This and definition of $p(\theta^*)$ together imply that such ϵ_n exists.

Combining (B.5) and (B.6) I get that

$$T_{n,0} = \sup_{\theta_0 \in \Theta_0 \bigcap \mathrm{NE}} \left\| \sqrt{n} \mu_n \left\{ g_n(\theta_{0n}) \right\} \right\|_e^2 + o_p(1).$$

Step 3. Define $\mathcal{G}_n = \{g(\theta_{0n}) : \theta_0 \in \Theta_0 \bigcap \mathrm{NE}\}$. Note that $\mathcal{G}_n$ is class of functions indexed by $n \in \mathbb{N}$ and θ_0 from a compact with respect to $\|\cdot\|_c$ set $\Theta_0 \bigcap \mathrm{NE}$. Since $\mathcal{G}_n \subseteq \mathcal{G}$ for all n, there exists in envelope g such that $\|g\|_{L^2(P_0)} = O(1)$ and $\|g \mathbb{1}(g > c\sqrt{n})\|_{L^2(P_0)} = o(1)$ for every $c > 0$. Moreover, by Corollary 2.7.2. in Van Der Vaart and Wellner (1996), for every $\alpha_n = o(1)$

$$\int_0^{\alpha_n} \sqrt{\log N_{[]} \left(\epsilon \|g\|_{L^2(P_0)}, \mathcal{G}_n, L^2(P_0) \right)} d\epsilon \leq \int_0^{\alpha_n} K \epsilon^{-\frac{d_x}{\kappa + \kappa_0}} d\epsilon = o(1).$$

Hence, if I show that for every $\alpha_n = o(1)$

$$\sup_{\|\theta_0 - \theta_0'\|_c < \alpha_n} \|g_n(\theta_{0n}) - g_n(\theta_{0n}')\|_{L^2(P_0)} = o(1),$$

then, by Theorem 2.11.23. in Van Der Vaart and Wellner (1996), the sequence

$$\left\{ \sqrt{n} \mu_n \left\{ g_n(\theta_{0n}) \right\}, \theta_0 \in \Theta_0 \bigcap \mathrm{NE} \right\}$$

is asymptotically tight in $l^\infty(\Theta_0 \bigcap \mathrm{NE})$ and converges in distribution to a tight Gaussian process provided the sequence of covariance functions $\langle g_n(\theta_{0n}), g_n(\theta_{0n}') \rangle_{L^2(P_0)}$ converges pointwise on $\Theta_0 \bigcap \mathrm{NE} \times \Theta_0 \bigcap \mathrm{NE}$.

Note that since $g_n(\theta_{0n}) \in \mathcal{G}$, there exists at least $[d_x/2]$-times continuously differentiable function $g(\theta_0)$ such that $\|g_n(\theta_{0n}) - g(\theta_0)\|_{L^2(P_0)} = o(1)$. Hence, by the triangular inequality,

$$\begin{aligned} &\sup_{\|\theta_0 - \theta_0'\|_c < \alpha_n} \|g_n(\theta_{0n}) - g_n(\theta_{0n}')\|_{L^2(P_0)} \leq \\ &\leq \sup_{\|\theta_0 - \theta_0'\|_c < \alpha_n} \|g(\theta_0) - g(\theta_0')\|_{L^2(P_0)} + o(1) \leq O(\alpha_n). \end{aligned}$$

B.6. Proof of Theorem 2.4

Let $\hat{\Theta}_n^w \subseteq \Theta_{k(n)}$ and $\tilde{\Theta}_n^w \subseteq \Theta_{k(n)}$ be collections of $\hat{\theta}_n^w = (\hat{\beta}_n^w, \hat{h}_n^w)$ and $\tilde{\theta}_n^w = (\tilde{\beta}_n^w, \tilde{h}_n^w)$ respectively that satisfy

$$\begin{aligned} L_n^w(\hat{\theta}_n^w) &= \sup_{\theta \in \Theta_{k(n)}} L_n^w(\theta), \\ L_n^w(\tilde{\theta}_n) &= \sup_{\theta \in \Theta_{k(n)} \bigcap \mathrm{NE}(\hat{\theta}_n)} L_n^w(\theta). \end{aligned}$$

First, I state the bootstrap versions of Lemmas presented in Appendix B.4.

Lemma B.8 *Under Assumptions 1, 2 and 3, if* $\delta_n \log\log(n) = o(1)$*, then uniformly in* $\theta_0 \in \Theta_0$*,* $\theta \in \mathcal{B}_n(\theta_0)$*,* $g \in \bar{\mathcal{G}}$ *and* $t_n \in \mathcal{T}_n$

$$\begin{aligned} &n^{-1}\left(L_n^w(\theta(t_n, g)) - L_n^w(\theta)\right) = \\ &t_n \mu_n\left\{wg\right\} - \frac{\|t_n\|_e^2}{2} - t_n \left\langle \frac{p(\theta)}{p_0} - 1, g \right\rangle_{L^2(P_0)} (1 + o(1)) + o_p(n^{-1}). \end{aligned}$$

Lemma B.9 *Under Assumptions 1, 2 and 3, uniformly in* $\theta_0 \in \Theta_0$

$$\sup_{\hat{\theta}_n^w \in \hat{\Theta}_n} \sup_{g \in \bar{\mathcal{G}}} \left\| \left\langle \frac{p(\hat{\theta}_n^w)}{p_0} - 1, g \right\rangle_{L^2(P_0)} (1 + o(1)) - \mu_n\left\{wg\right\} \right\|_e = o_p(n^{-1/2}).$$

Lemma B.10 *If* $\gamma_n = o(n^{-1/2})$*, then for every* $\theta_0 \in \Theta_0$ *such that* $\hat{\theta}_n \in B_n(\theta_0)$*,* $j = 0, 1, 2, 3$*, there exists* $g_n(\theta_{0n}) \in \bar{\mathcal{G}}$ *such that*

$$\sup_{\theta \in B_n(\theta_0) \bigcap \mathrm{NE}_j(\hat{\theta}_n)} \left| \left\langle \frac{p(\theta)}{p_0} - \frac{p(\hat{\theta}_n)}{p_0}, g_n(\theta_{0n}) \right\rangle_{L^2(P_0)} \right| = o_p(n^{-1/2}).$$

Second, I mimic the steps of the proof of Theorem 2.3.

Without loss of generality I consider the case when only the equality constraint is binding. The proof consists of three steps.

Step1. Take an arbitrary $\theta_0 \in \Theta_0$ and $\theta \in B_n(\theta_0) \bigcap \mathrm{NE}(\hat{\theta}_n)$. Consider a perturbation in the probability density sieve space: for $g_n(\theta_{0n})$ from Lemma B.3

$$p(\theta(t_n, g_n(\theta_{0n}))) = p(\theta) + t_n g_n(\theta_{0n}) p_0.$$

By Lemmas B.8 and B.10

$$n^{-1}[L_n^w(\theta_n(t_n, g_n(\theta_{0n}))) - L_n^w(\theta_n)] = t_n\mu_n\left\{(w-1)g_n(\theta_{0n})\right\} - \frac{t_n^2}{2} + o_p(n^{-1})$$

Then by Lemma B.4 and properties of the quadratic forms, the maximizer belongs to $\mathcal{T}_n$ and is equal to $\mu_n\left\{(w-1)g_n(\theta_{0n})\right\}$. Hence,

$$\sup_{t_n\in\mathcal{T}_n} n^{-1}[L_n^w(\theta(t_n, g_n(\theta_{0n},\beta))) - L_n^w(\theta)] - o_p(n^{-1}) = \frac{\left\|\mu_n\left\{(w-1)g_n(\theta_{0n})\right\}\right\|_e^2}{2} + o_p(n^{-1})$$

As a result,

$$\begin{aligned}
&n^{-1}[L_n^w(\hat{\theta}_n^w) - L_n^w(\tilde{\theta}_n^w)] \geq \sup_{\theta_0\in\Theta_0}\inf_{\theta\in B_n(\theta_0)} n^{-1}[L_n^w(\hat{\theta}_n^w) - L_n(\theta)] + o_p(n^{-1}) \geq \\
&\sup_{\theta_0\in\Theta_0}\inf_{\theta\in B_n(\theta_0)}\sup_{t_n\in\mathcal{T}_n} n^{-1}[L_n^w(\theta(t_n, g_n(\theta_{0n}))) - L_n(\theta)] + o_p(n^{-1}) = \\
&\sup_{\theta_0\in\Theta_0}\frac{\left\|\mu_n\left\{(w-1)g_n(\theta_{0n})\right\}\right\|_e^2}{2} + o_p(n^{-1}).
\end{aligned}$$

Finally,

$$n^{-1}[L_n^w(\hat{\theta}_n^w) - L_n(\tilde{\theta}_n^w) \geq \sup_{\theta_0\in\Theta_0}\frac{\left\|\mu_n\left\{(w-1)g_n(\theta_{0n})\right\}\right\|_e^2}{2} + o_p(n^{-1}) \qquad \text{(B.7)}$$

Step 2. Consider the following perturbation of $\hat{\theta}_n^w$: t_n^* and $g_n(\theta_{0n})$ such that

$$\begin{aligned}
t_n^* &= -\left\langle\frac{p(\hat{\theta}_n^w)}{p_0} - \frac{p(\hat{\theta}_n)}{p_0}, g_n(\theta_{0n})\right\rangle_{L^2(P_0)} + \epsilon_n, \\
p(\theta^*) &= p(\hat{\theta}_n^w) + t_n^* g_n(\theta_{0n})p_0,
\end{aligned}$$

where $\epsilon_n = o_p(n^{-1/2})$. Assume for a moment that for some ϵ_n, $\theta^* \in B_n(\theta_0)\bigcap \mathrm{NE}(\hat{\theta}_n)$. Then, since $t_n^* = O_p(n^{-1/2})$ (Lemma B.7 and Lemma B.4) and by applying Lemma B.8 and Lemma B.7 to t^* and $g_n(\theta_{0n})$, one can get that

$$n^{-1}[L_n^w(\hat{\theta}_n^w) - L_n^w(\tilde{\theta}_n^w)] \leq n^{-1}[L_n^w(\hat{\theta}_n^w) - L_n^w(\theta^*)] \leq \frac{\|t_n^*\|_e^2}{2} + o_p(n^{-1})$$

After applying Lemma B.7 to t^* and $g_n(\theta_0)$, I get that

$$n^{-1}[L_n^w(\hat{\theta}_n) - L_n^w(\tilde{\theta}_n)] \leq \sup_{\theta_0\in\Theta_0}\frac{\|\mu_n\{(w-1)g_n(\theta_{0n})\}\|_e^2}{2} + o_p(n^{-1}). \qquad \text{(B.8)}$$

It remains to show that such ϵ_n exists. Note that

$$\left\langle \frac{p(\theta^*)}{p_0} - \frac{p(\hat{\theta}_n)}{p_0}, g_n(\theta_{0n}) \right\rangle_{L^2(P_0)} = - \left\langle \frac{p(\hat{\theta}_n^w)}{p_0} - \frac{p(\hat{\theta}_n)}{p_0}, g_n(\theta_{0n}) \right\rangle_{L^2(P_0)} +$$
$$+ \left\langle \frac{p(\hat{\theta}_n^w)}{p_0} - \frac{p(\hat{\theta}_n)}{p_0}, g_n(\theta_{0n}) \right\rangle_{L^2(P_0)} + \epsilon_n = \epsilon_n = o_p(n^{-1/2})$$

This and definition of $p(\theta^*)$ together imply that such an ϵ_n exists.

Combining (B.7) and (B.8) I get that

$$T_{n,0}^w = \sup_{\theta_0 \in \Theta_0} \left\| \sqrt{n} \mu_n \left\{ (w-1) g_n(\theta_{0n}) \right\} \right\|_e^2 + o_p(1).$$

Step 3. This step is essentially coincides with Step 3 in the proof of Theorem 2.3.

By Theorem 2.9.7 in Van Der Vaart and Wellner (1996),

$$T_{n,0}^w \Rightarrow \sup_{g \in \mathcal{G}} \|G(g)\|_e^2 \quad \text{a.s.},$$

where $\mathcal{G}$ is a set of $L^2(P_0)$ limit points of $\{g_n(\theta_{0n})\}_{\theta_0 \in \Theta_0}$, $G(\cdot)$ is a centered Gaussian process with the covariance function $\mathbb{E}[\, gg' \,]$.

The result then follows from the fact that in the bootstrap statistic supremum is taken over a bigger set. Hence, the critical values of the bootstrap statistic are weakly bigger than the corresponding critical values of the original statistic with probability 1.

Appendix C

Omitted proofs for Chapter 3

Proof of Proposition 2. Conditions (i) and (ii) allow us to use Claim 1. Suppose $\xi = z$ and without loss of generality assume that $F(\cdot)$ is a strictly increasing function. Fix some $\theta \in \Theta$. Denote

$$\alpha(z,\theta) = \frac{qF'(z^T\theta)}{G(z,\theta)^2}$$
$$A = \mathbb{E}\left[\,\alpha(z,\theta)\mathbf{z}\mathbf{z}^T\,\right]$$

Then

$$-\mathbb{E}[\,\partial_{\theta^T} m(\mathbf{z},\theta)\,] = A$$

A is positive definite since $\inf_z \alpha(z,\theta) > 0$ for all θ. The case when $\xi = x$ can be proved analogously by replacing z by x and $\alpha(z,\theta)$ by $\gamma(x,\theta) = qF'(z^T\theta)$.

■

Bibliography

Andrews, D. W., Berry, S., and Jia, P. (2004). Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location. *manuscript, Yale University*.

Andrews, D. W. and Shi, X. (2013). Inference based on conditional moment inequalities. *Econometrica*, 81(2):609–666.

Andrews, D. W. K. (2011). Examples of L2-complete and boundedly-complete distributions. Discussion Paper 1801, Cowles Foundation.

Aradillas-López, A. (2010). Semiparametric estimation of a simultaneous game with incomplete information. *Journal of Econometrics*, 157(2):409–431.

Aradillas-López, A. and Rosen, A. (2013). Inference in ordered response games with complete information. Working Paper CWP33/13, CeMMAP.

Aradillas-López, A. and Tamer, E. (2008). The identification power of equilibrium in simple games. *Journal of Business & Economic Statistics*, 26(3):261–283.

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

Bajari, P., Hong, H., and Ryan, S. P. (2010b). Identification and estimation of a discrete game of complete information. *Econometrica*, 78(5):1529–1568.

Battalio, R., Samuelson, L., and Huyck, J. V. (2001). Optimization incentives and coordination failure in laboratory stag hunt games. *Econometrica*, 69(3):749–764.

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

Berry, S. T. (1990). Airport presence as product differentiation. *The American Economic Review*, pages 394–399.

Berry, S. T. and Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797.

Berry, S. T., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Berry, S. T. and Tamer, E. (2006). Identification in models of oligopoly entry. In Blundell, R., Newey, W. K., and Persson, T., editors, *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress, Volume 2*, volume 42 of *Econometric Society Monographs*, pages 46–85. Cambridge University Press.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*, 35(7):677–693.

Bjorn, P. A. and Vuong, Q. H. (1984). Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation. Social Science Working Paper 537, California Institute of Technology.

Bresnahan, T. F. and Reiss, P. C. (1990). Entry in monopoly market. *The Review of Economic Studies*, 57(4):531–553.

Bresnahan, T. F. and Reiss, P. C. (1991). Empirical models of discrete games. *Journal of Econometrics*, 48(1–2):57–81.

Brown, L. D. (1986). *Fundamentals of statistical exponential families with applications in statistical decision theory*, volume 9 of *Lecture notes – Monograph series*. Institute of Mathematical Statistics.

Camerer, C. (2003). *Behavioral Game Theory: experiments in strategic interaction*. Princeton University Press.

Canay, I. A., Santos, A., and Shaikh, A. M. (2013). On the testability of identification in some nonparametric models with endogeneity. *Econometrica*, 81(6):2535–2559.

Carlsson, H. and Van Damme, E. (1993). Global games and equilibrium selection. *Econometrica*, 61(5):989–1018.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334.

Chen, X. and Liao, Z. (2014). Sieve m inference on irregular parameters. *Journal of Econometrics*, 182(1):70–86.

Chen, X. and Pouzo, D. (2015). Sieve quasi likelihood ratio inference on semi/nonparametric conditional moment models. *Econometrica, forthcoming.*

Chen, X., Tamer, E., and Torgovitsky, A. (2011). Sensitivity analysis in semiparametric likelihood models. *Yale University and Northwestern University.*

Chernozhukov, V., Newey, W. K., and Santos, A. (2015). Constrained conditional moment restriction models. *arXiv preprint arXiv:1509.06311.*

Chiappori, P.-A., Levitt, S., and Groseclose, T. (2002). Testing mixed-strategy equilibria when players are heterogeneous: the case of penalty kicks in soccer. *American Economic Review*, 92(4):1138–1151.

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

Cosslett, S. (1981a). Efficient estimation of discrete-choice models. *Structural analysis of discrete data with econometric applications*, pages 51–111.

Cosslett, S. (1981b). Maximum likelihood estimator for choice-based samples. *Econometrica*, pages 1289–1316.

Crawford, V. P., Costa-Gomes, M. A., and Iriberri, N. (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*, 51(1):5–62.

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

Epstein, L. G., Kaido, H., and Seo, K. (2015). Robust confidence regions for incomplete models. *Econometrica, forthcoming.*

Florens, J.-P. (2003). Inverse problems and structural econometrics. In Dewatripont, M., Hansen, L. P., and Turnovsky, S. J., editors, *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Volume 2*, volume 34 of *Econometric Society Monographs*, pages 46–85. Cambridge University Press.

Fox, J. T. (2007). Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *The RAND Journal of Economics*, 38(4):1002–1019.

Freyberger, J. (2014). Testing completeness and weak instruments in nonparametric instrumental variable models. Mimeo.

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

Graham, C., Ferrier, S., Huettman, F., Moritz, C., and Peterson, A. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, 19(9):497–503.

Green, E. J., Marshall, R. C., and Marx, L. M. (2013). Tacit collusion in oligopoly. Mimeo.

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

Hall, P. and Horowitz, J. L. (2005). Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, 33(6):2904–2929.

Halpern, J. Y. and Pass, R. (2012). Game theory with translucent players. Cornell Univeristy, Mimeo.

Harsanyi, J. C. and Selten, R. (1988). *A general theory of equilibrium selection in games*. The MIT Press.

Hellerstein, J. and Imbens, G. (1999). Imposing moment restrictions from auxiliary data by weighting. *Review of Economics and Statistics*, 81(1):1–14.

Henry, M. and Mourifie, I. (2012). Nonparametric sharp bounds for payoff functions in 2x2 games. Mimeo.

Hirano, K., Imbens, G., Ridder, G., and Rebin, D. (1998). Combining panel data sets with attrition and refreshment samples.

Hoderlein, S., Nesheim, L., and Simoni, A. (2012). Semiparametric estimation of random coefficients in structural economic models. Working Paper CWP09/12, CeMMAP.

Hölmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 13(2):324–340.

Hsu, Y.-C. and Shi, X. (2013). Model selection tests for conditional moment inequality models. Technical report, Institute of Economics, Academia Sinica, Taipei, Taiwan.

Imbens, G. (1992). An efficient method of moments estimator for discrete choice models with choice-based sampling. *Econometrica*, pages 1187–1214.

Kashaev, N. and Salcedo, B. (2016). Identification of solution concepts for complete information games. Working paper.

Kitamura, Y. (2000). Comparing misspecified dynamic econometric models using nonparametric likelihood. *Department of Economics, University of Wisconsin*.

Kline, B. (2015a). The empirical content of games with bounded regressors. *Quantitative Economics, forthcoming*.

Kline, B. (2015b). An empirical model of non-equilibrium behavior in games. Working paper.

Kline, B. and Tamer, E. (2012). Bounds for best response functions in binary games. *Journal of Econometrics*, 166(1):92–105.

Lancaster, T. and Imbens, G. (1996). Case-control studies with contaminated controls. *Journal of Econometrics*, 71(1):145–160.

Levitt, S. and Porter, J. (2001). How dangerous are drinking drivers? *Journal of Political Economy*, 109(6):1198–1237.

Liu, X. and Shao, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *Annals of Statistics*, pages 807–832.

Manski, C. and Lerman, S. (1977). The estimation of choice probabilities from choice based samples. *Econometrica*, pages 1977–1988.

Molchanov, I. (2006). *Theory of random sets*. Springer.

Morris, S. and Shin, H. S. (2003). Global games: theory and applications. In Dewatripont, M., Hansen, L. P., and Turnovsky, S. J., editors, *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Volume 1*, volume 34 of *Econometric Society Monographs*, pages 56–114. Cambridge University Press.

Murphy, S. A. and Van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–465.

Nevo, A. (2003). Using weights to adjust for sample selection when auxiliary information is available. *Journal of Business & Economic Statistics*, 21(1):43–52.

Newey, W. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578.

Palacios-Huerta, I. (2003). Professionals play minimax. *The Review of Economic Studies*, 70(2):395–415.

Pearce, J. and Boyce, M. (2005). Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, 43(3):405–412.

Ridder, G. (1992). An empirical evaluation of some models for non-random attrition in panel data. *Structural Change and Economic Dynamics*, 3(2):337–355.

Rivers, D. and Vuong, Q. (2002). Model selection tests for nonlinear dynamic models. *The Econometrics Journal*, 5(1):1–39.

Rubinstein, A. (1989). The electronic mail game: strategic behavior under almost common knowledge. *American Economic Review*, 79(3):385–391.

Salcedo, B. (2012). Implementation without commitment in moral hazard environments. Pennsylvania State University, Mimeo.

Santos, A. (2012). Inference in nonparametric instrumental variables with partial identification. *Econometrica*, 80(1):213–275.

Shen, X. and Shi, J. (2005). Sieve likelihood ratio inference on general parameter space. *Science in China Series A: Mathematics*, 48(1):67–78.

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

Tao, J. (2014). Inference for point and partially identified semi-nonparametric conditional moment models. Technical report, Working paper, University of Wisconsin-Madison.

Tripathi, G. (2011). Gmm based inference with stratified samples when the aggregate shares are known. *Journal of Econometrics*.

Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence*. Springer.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307–333.

Vita
Nail Kashaev

Education

- *Ph.D. in Economics* — Pennsylvania State University, 2016
  Fields: Econometric Theory, Industrial Organization, Applied Econometrics
  Committee: Joris Pinkse, Sung Jae Jun, Andrés Aradillas-López, Edward Jaenicke
- *M.A. in Economics* — New Economic School, 2010
- *M.S. in Applied Mathematics and Physics* — Moscow Institute of Physics and Technology, 2010
- *B.S. in Applied Mathematics and Physics* — Moscow Institute of Physics and Technology, 2008

Papers

- *Testing for Nash behavior in entry games with complete information* (2016), Job Market Paper
- *Identification of solution concepts for discrete semi-parametric games with complete information* (2016), with Bruno Salcedo, working paper
- *Parametric estimation of binary models with contaminated controls* (2013), working paper

Conferences & Seminars

- *Testing for Nash behavior in entry games with complete information* — Presented at Cornell University, Western University, University of Oslo, Toulouse School of Economics, Higher School of Economics (Moscow)
- *Identification of solution concepts for discrete semi-parametric games of complete information* — Joint work with Bruno Salcedo, Presented at Midwest Theory Spring Conference (Indianapolis, 2014), NAMES (Minnesota, 2014)
- *Parametric estimation of binary models with contaminated controls* — Presented at Midwest Econometrics Group Meeting (Bloomington, 2013), LAMES (Colmex, 2013)

Teaching experience

- *Teaching assistant* — Introductory Microeconomic Analysis and Policy (PSU, 2010), Intermediate Macroeconomics (PSU, 2011) Intro Econometrics, (PSU, 2012) Money and Banking, (PSU, 2013)
- *Teaching assistant for Ph.D. core courses* — Econometrics (PSU, 2012, 2014, 2015)

Fellowships & Awards

- Grow Award (PSU, 2014), Abramov Scholarship (MIPT, 2006-2008) Mirakov Scholarship (MIPT, 2007)