Testing for Nash behavior in binary games of complete information*

Nail Kashaev†
nkashaev@uwo.ca

December 1, 2017

Abstract *Nash equilibrium* (NE) is a leading solution concept in the empirical analysis of games. The assumption of NE play is crucial, not just for estimation, but also for the validity of counterfactual exercises and policy implications. I propose a sieve-likelihood-ratio procedure to test the NE assumption in a complete-information binary game with *level-2 rational* players. The method is robust to partial identification and allows for the nonparametric selection of equilibria. The proposed procedure is applied to data on entry decisions of small grocery stores in rural areas in the USA, as used in Grieco (2014).

Keywords: Model selection, econometrics of games, multiple equilibria, partial identification, sieve estimation

*I am thankful for the supervision of Joris Pinkse and Sung Jae Jun, as well as the useful comments from Andrés Aradillas-López, Victor H. Aguiar, Victor Aguirregabiria, Roy Allen, Tim Conley, Russell Cooper, Ronald Gallant, Sílvia Gonçalves, Paul Grieco, Patrik Guggenberger, Martin Hackmann, Marc Henry, Yongmiao Hong, Nicholas Kiefer, Kala Krishna, Vijay Krishna, Rohit Lamba, Huihui Li, Rosa Matzkin, Francesca Molinari, Marcelo Moreira, Salvador Navarro, Bruno Salcedo, Jörg Stoye, Bruno Sultanum, Neil Wallace, and seminar participants at Cornell University, Higher School of Economics (Moscow), Penn State, Toulouse School of Economics, University of Calgary, University of Oslo, University of Toronto, University of Western Ontario. I also thank the attendants of the 2016 Midwest Econometrics Group Meeting, the 2016 Summer Meeting of The Econometric Society, and the 2017 Canadian Econometric Study Group Meeting.

†Department of Economics, University of Western Ontario.

1. Introduction

This paper considers the problem of testing whether the assumption of *Nash equilibrium* (NE) is consistent with observed data in semiparametric binary games with complete information. The assumption maintained is the *level-2 rationality* of players. That is, I assume that every player is a utility maximizer and knows that her opponents are also maximizing their utilities. I propose a testing procedure that does not assume point identification of parameters of the model and does not impose parametric restrictions on the way players randomize among different equilibria in regions where an equlilibrium is not unique.

My procedure consist of three steps. First, under level-2 rationality, I describe the distribution over outcomes conditional on covariates as a function of a finite-dimensional parameter governing the distribution of payoffs and an infinite-dimensional parameter describing the behavior of players. I call this infinite-dimensional parameter *the distribution of play*. The distribution of play is a conditional distribution over the set of possible outcomes of the game conditional on a realization of observed and unobserved payoff shifters. Second, I show that the NE assumption can be characterized by a finite set of equality and inequality constraints on parameters of the model. Third, I construct a sieve-likelihood-ratio (sLR) test statistic by taking the difference between unconstrained (implied by level-2 rationality) and constrained (implied by NE) log-likelihood objective functions.

Testing for NE behavior is important for several reasons. Testing the NE assumption against level-2 rationality is equivalent to testing whether players' beliefs are consistent. Second, many estimation methods are based on the NE assumption. Thus, a violation of NE behavior may lead to inconsistent estimates of the payoff parameters. Third, even if one employs estimation methods that are robust to the failure of the NE assumption, the counterfactual predictions based on those robust estimates can be either wrong (if one incorrectly assumes NE behavior) or uninformative (if one assumes, for instance, only level-2 rationality). Finally, while the validity of different equilibrium concepts has been studied in the experimental game theory literature (e.g., Camerer (2003), Crawford et al. (2013), Kline (2015b)), the correctness of the NE assumption has not been fully addressed in settings where payoffs are unknown. This paper aims to fill this gap.

The level-2 rationality assumption is essential for my procedure. It is general enough to allow agents to play any strategy that survives two rounds of strict dominance elimination, e.g., a Nash or correlated equilibrium. However, the agents are not allowed to play maximin strategies.[1] For entry games the NE assumption is equivalent to assuming (i) the level-2 rationality of players and (ii) the consistency of each player's beliefs about the actions of opponents. By maintaining the level-2 rationality assumption, I am implicitly testing whether players have consistent beliefs about each other's actions. For instance, if market conditions are such that only a monopolist can make positive profits, then under level-2 rationality it is possible that no one decides to enter. Players may believe that at least one competitor will enter the market. On the other hand, under the NE assumption there should be at least one entrant with positive probability; for example, firm 1 believes that no one else is entering, and other firms believe that firm 1 enters.

I work with a likelihood-based model instead of a model based on conditional moment inequalities for several reasons. First, in the point-identified case it is efficient. Second, in contrast to general model specification tests based on moment inequalities, a likelihood-based model allows me to test consistency of beliefs. Thus, my testing procedure directs power against a specific set of alternatives. Third, my procedure does not require the researcher to choose tuning parameters. Finally, in practice, the researcher often needs to compute confidence sets for payoff parameters or the distribution of play. With slight modification of my test statistic one can construct confidence sets for parameters of interest using the methodology proposed in Chen et al. (2011).[2]

I emphasize that the level-2 rationality of players and the parametric utility specification are assumed to hold under both the null hypothesis and the alternative one. Asymptotic behavior of the test statistics is unknown if either one of these assumptions is violated. Thus, any conclusion based on my testing procedure is valid only if both the correct parametric specification of the payoff distribution and level-2 rationality are assumed. In other words, my testing procedure is a diagnostic tool that can help the researcher to assess the validity of the NE assumption when the parametric specification of the payoffs and rationality of players are not in question.

My approach to modeling binary games, which emphasizes the notion of the distribution of play, is different from the existing literature in several respects. First, the standard approach is to assume a particular equilibrium concept, such as NE, and then augment the model with a selection mechanism (distribution over equilibria). [3] 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 does not depend on any notion of equilibrium behavior because it is the distribution over possible outcomes (not over equilibria). Hence, I parametrize the model independently

[1] For a detailed discussion about level-2 rationality, see Aradillas-López and Tamer (2008).

[2] See section 8 for more details.

[3] For instance, Bajari et al. (2010) use a parametric selection mechanism.

of equilibrium concepts. In my setting any equilibrium concept is just a set of restrictions on parameters. Although I am focusing on testing the NE assumption, I can test any equilibrium restrictions as long as these restrictions are consistent with the level-2 rationality assumption and can be characterized by a set of equality/inequality constraints that are smooth in the payoff parameters and affine in the distribution of play (e.g., a correlated equilibrium). Second, I do not treat the distribution of play as a 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 infer whether agents play NE only in pure strategies or whether they play the mixed-strategy NE as well.

My testing procedure is based on the maximum likelihood approach with finite-dimensional sieves. Hence, it inherits all its advantages. Once the infinite-dimensional parameter is replaced with its sieve approximation, the problem becomes parametric for implementation purposes, and is analogous to parametric maximum likelihood estimation. Thus, the procedure is computationally tractable. Moreover, since my procedure is based on the likelihood function I do not need to estimate the identified set. Although the parameters of the model may be set-identified, the density implied by these parameters is point-identified.

The empirical literature on parametric complete information games has focused on the estimation of payoff parameters under the NE assumption.[4] A substantial part of this literature has analyzed binary games of the kind studied in entry models. For instance, Ciliberto and Tamer (2009) analyze the market structure of the US airline industry assuming NE play in pure strategies.[5] That is why, for exposition purposes, I focus on a two-player entry game. However, I show how the proposed methodology can be extended to coordination games and to binary games with more than two players.

I provide Monte Carlo evidence of the finite-sample performance of my procedure. I also compare the rejection probabilities of my method to the rejection probabilities of alternative procedure based on Beresteanu et al. (2011). In simulation studies my method delivers less conservative results than the method based on Beresteanu et al. (2011). I then apply my testing procedure to the complete information version of the entry model presented in Grieco (2014).

[4] Relevant examples include papers about discrete complete-information games: Bjorn and Vuong (1984), Berry (1990), Bresnahan and Reiss (1990, 1991), Tamer (2003), Berry and Tamer (2006), Aradillas-López and Tamer (2008), Ciliberto and Tamer (2009), Bajari et al. (2010), Beresteanu et al. (2011), Galichon and Henry (2011), Henry and Mourifie (2012), Kline and Tamer (2012), Aradillas-López and Rosen (2013), Kline (2015a).

[5] Their procedure can be extended to the case in which players are allowed to play mixed-strategy NE.

Related Literature

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 confidence sets which are never empty by construction. Andrews et al. (2004), Ciliberto and Tamer (2009), Aradillas-López and Rosen (2013), and Epstein et al. (2015), among others, have developed techniques to construct confidence sets for 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, 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 a 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 and with nonparametric selection mechanisms 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)* there is a potential loss in power of by-product tests against my alternatives since these tests are general model specification tests; *(iii)* under my setting it is easy to impose additional restrictions on the distribution of play, such as exclusion and/or monotonicity restrictions.

Bugni et al. (2015) propose a model specification test that outperforms the by-product test of Andrews and Soares (2010). However, their procedure can only be applied to models characterized by *unconditional* moment inequalities. Thus, the NE assumption can be tested within their framework only if covariates are discrete.

The paper is closely related to the literatures on (semi)parametric LR or quasi-LR (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 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.

Despite their similarity to my approach, the main results in CTT and Tao (2014) cannot be applied to my problem. First, since the NE constraints are infinite-dimensional, I need to aggregate them without losing information. Second, there is no one-to-one mapping between the NE constraints and residual functions (Assumption 3.4 in Tao (2014)). Third, I cannot use the L^2-inner product and the norm induced by that inner product, which are presented in CTT, to linearize the NE constraints. Finally, since the linear approximation of the constraints can be applied only under the null hypothesis, the multiplier bootstrap procedure I am proposing will not approximate quantiles of the asymptotic null distribution of the test statistic. Instead, it will approximate quantiles of the distribution that dominates the distribution of interest. This is not the case for the multiplier bootstrap procedure used in CTT and Tao (2014).

Notation and definitions

I use boldface font to denote random objects and regular font for deterministic ones. For a column vector $\beta \in \mathbb{R}^{d_\beta}$, $\beta^\intercal$, β_i, β_{-i}, and $\|\beta\|_e$ denote its transpose, the i-th component, $(\beta_1, \ldots, \beta_{i-1}, \beta_{i+1}, \ldots, \beta_{d_\beta})^\intercal$, 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 $(f(y_1), f(y_2), \ldots, f(y_k))^\intercal$. For $f : X \times \Theta \to \mathbb{R}$, where $X \subset \mathbb{R}^{d_x}$, let $\partial_x f(x_0, \theta_0)$ denote the vector of partial derivatives of $f(\cdot, \theta_0)$ evaluated at x_0; $f(\theta_0) = f(\cdot, \theta_0)$ and $\partial_{x^\intercal} f(x_0, \theta_0) = (\partial_x f(x_0, \theta_0))^\intercal$. 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 those of $\mathbf{x}_1$ conditional on $\mathbf{x}_2 = x_2$, respectively. For $d \in \mathbb{R}$, let $[d]$ denote the integer part of d. For $\{\mathbf{x}_n\}_{n=1}^\infty$ and $\{\alpha_n\}_{n=1}^\infty$, the relations $\mathbf{x}_n = O_p(\alpha_n)$ 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.

Structure of the Paper

The paper is organized as follows. Section 2 describes the setting of the game and assumptions on observables and unobservables. Section 3 describes the constraints imposed on the model by the level-2 rationality and NE assumptions in the two-player entry game. In section 4 I describe assumptions on the parameter space and define the identified set. Section 5 provides results on consistency and the rate of convergence of the sieve maximum likelihood estimator in terms of the Pearson distance. Section 6 describes the testing procedure and the null distribution of statistics involved. I provide results on the multiplier bootstrap in Section 7. Section 8 shows how the proposed procedure can be extended to coordination games and binary games with more than two players. In section 9 I provide results for Monte Carlo experiments and an empirical application based on the model and data presented in Grieco (2014). Section 10 concludes.

2. An entry game

I start by considering a two-player entry game. Later on, in section 8, I show how my procedure can be applied to coordination games and to binary games with more than two players.

2.1. Payoff functions

Consider a two-player entry game with complete information. Two firms $i \in I = \{1, 2\}$ must simultaneously decide whether to enter a market ($\mathbf{y}_i = 1$) or not ($\mathbf{y}_i = 0$). 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:

$$u_i(y, x, e, \beta) = (\bar{x}_i^\intercal \bar{\beta}_i - \tilde{x}_i^\intercal \tilde{\beta}_i y_{-i} - e_i) y_i, \tag{1}$$

where $\mathbf{x} \in X \subseteq \mathbb{R}^{d_x}$ is the vector of covariates that is composed of the nonredundant elements of $\bar{x}_i$ and $\tilde{x}_i$, $i \in I$, and is observed by both players and the econometrician. The part of the payoffs unobserved by the econometrician is $\mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2)^\intercal$. 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 $\bar{x}_i^\intercal \bar{\beta}_i - e_i$ represents i's benefit or cost of entering the market, while $\tilde{x}_i^\intercal \tilde{\beta}_i$ 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 $\bar{\beta}_i$, $\tilde{\beta}_i$,

$i \in I$, and the parameters of the conditional distribution of unobservables.

For the sake of exposition, I assume the linear specification of payoffs. 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 in 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 $\mathbf{z} \in \{\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2\}$.

Unobservables:

(iv) $\mathbf{e}|\mathbf{x} = x$ 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.

(v) $f_{e|x}$ 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 $X \times B$.

(vi) $f_{e|x}$ 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 and can be relaxed to some extent. For instance, I can allow for covariates with discrete support (as I do in my empirical application) by introducing sieve spaces for each point in their finite support. However, compactness of X can not be relaxed.[6] Assumption 1.(iii) is a rank condition for index models. Assumption 1.(iv) is standard in the literature and guarantees that every outcome is realized with positive probability conditional on the realization of covariates. Assumptions 1.(v) and 1.(vi) are satisfied by many parametric distributions. For instance, one can assume that $(\mathbf{e}_1, \mathbf{e}_2)^\mathsf{T}$ are independent of $\mathbf{x}$ and jointly normal with the unknown correlation parameter ρ.[7]

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

[6] If one assumes that there is a covariate with unbounded support, then there are no economic restrictions that can control the tail behaviour of the infinite-dimensional parameter. Thus, one can not ensure that the sieve approximation behaves well when $\|x\|$ takes arbitrary large values.

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

Assumption 2 is standard in the literature; it requires firms to prefer monopolies over duopolies. The signs of the competition effects determine whether one analyzes the entry game or the coordination game. Assumption 2 can be replaced by any set of assumptions that guarantee that the signs of the competition effects do not change with x and are known to the econometrician.

2.2. Distribution of play and equilibrium restrictions

The above model is incomplete, since I have not specified the equilibrium assumptions. Next, I describe the behavior of players using the notion of *the distribution of play* introduced in Kashaev and Salcedo (2016).

Definition 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 is a "reduced form" parameter. It is well-defined independently of the solution concept (e.g., Nash equilibrium or level-2 rationality). However, without additional structure there is no connection between payoffs and the distribution of play. Solution concepts, such as NE and level-2 rationality, build a bridge between the distribution of play and the payoff distribution.

Definition 2 Let U be the image of $u(\cdot, \cdot, \cdot) = (u_i(y, \cdot, \cdot, \cdot))_{i\in \mathrm{I}, y\in Y}$. Given a common probability space $(\Omega, \mathscr{F}, \Pr)$, a *solution concept* is a nonempty-valued correspondence $\mathrm{eq} : U \rightrightarrows \Delta(Y)$, such that

$$\left\{\omega \in \Omega \,\middle|\, \quad \Psi \cap \mathrm{eq}(u(\mathbf{x}(\omega), \mathbf{e}(\omega), \beta)) \neq \emptyset\right\} \in \mathscr{F}, \tag{2}$$

for every $\beta \in B$ and every *closed* set $\Psi \subseteq \Delta(Y)$, where $\Delta(Y)$ is a set of all distributions over Y.

Solution concepts determine the set of possible distribution over outcomes for a given realization of payoffs.[8]

[8]The definition of a solution concept can be extended to allow for incomplete-information solution concepts such as Bayesian-Nash equilibrium. One just needs to add beliefs as an argument of eq.

Definition 3 A pair $\theta = (\beta, h) \in \Theta$ is consistent with a solution concept eq if

$$h(\mathbf{x}, \mathbf{e}) \in \text{eq}(u(\mathbf{x}, \mathbf{e}, \beta)) \quad \text{a.s.}, \tag{3}$$

where $h(\cdot, \cdot) = (h(y, \cdot, \cdot))_{y \in Y}$.

Note that the solution-concept restrictions in (3) are deterministic. For every fixed $\theta = (\beta, h)$ one can always determine whether θ is consistent with any given eq. 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, or 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.[9] When the researcher has a particular equilibrium assumption 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 a selection mechanism in regions of multiplicity (see example 3.2).

Example 2.1 Suppose that the above entry game is a simultaneous move game of complete information; $\tilde{x}_i^{\intercal}\tilde{\beta}_i = \tilde{\beta}_i > 0$, $i \in I$. Assume also that firms always play a PNE and always play $y = (1, 0)$ whenever possible. It then follows that:

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

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

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

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, such h is not consistent with the NE assumption.

[9]See example 6.1 in Kashaev and Salcedo (2016).

3. Level-2 rationality 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 show how a relatively weak assumption on the behavior of players, namely level-2 rationality, can be used to construct a conditional density implied by a smooth h; and I construct a system of constraints on parameters that are equivalent to the NE assumption.

3.1. Level-2 rationality constraints

Assume that the firms are maximizing expected payoffs and have mutual knowledge of such behavior. For a 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. For example,

$$A((0,0), x, \beta) = \{e \in \mathbb{R}^2 \mid 0 > \bar{x}_i^\intercal \bar{\beta}_i - e_i,\ i = 1, 2\}.$$

In region $A((0,0), x, \beta)$ the best possible payoff that firms can get from entering (the monopoly payoff) is strictly smaller than zero. Hence, "not entering" is a strictly-dominant strategy for both firms. Similarly, in region $A((1,1), x, \beta)$ the worst possible payoff from entering (the duopoly payoff) is better than the payoff from $y_i = 0$. As a result, both firms will enter the market.

Let $A^M(x, \beta) = \mathbb{R}^2 \setminus \bigcup_{y \in Y} A(y, x, \beta)$ be the region in which the level-2 rationality assumption does not give unique predictions about the behavior of players. Figure 1 shows how these sets are constructed.

The level-2 rationality assumption can then be defined in terms of primitives of the model, as follows:

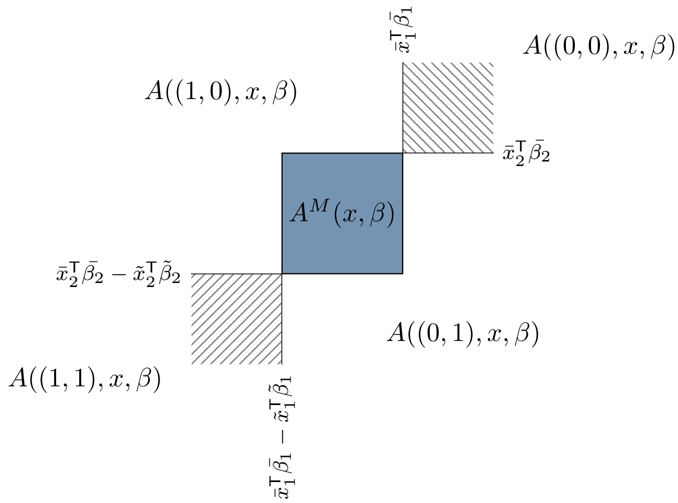


Figure 1 – Possible outcomes of the game under the level-2 rationality assumption.

Definition 4 A point $\theta \in \Theta$ is consistent with level-2 rationality if and only if

$$h(\mathbf{y}, \mathbf{x}, \mathbf{e}) \in \begin{cases} \{1\} & \text{if } \mathbf{e} \in A(\mathbf{y}, \mathbf{x}, \beta), \\ [0,1] & \text{if } \mathbf{e} \in A^M(\mathbf{x}, \beta), \\ \{0\} & \text{otherwise} \end{cases} \qquad \text{a.s.} \tag{4}$$

Assuming level-2 rationality substantially simplifies the model's predictions. We know exactly what the distribution of play should look like in the regions in which there is only one rationalizable outcome. The only region in which level-2 rationality does not impose any restrictions on the behavior of players is the region of multiplicity $A^M(x, \beta)$. In the region of multiplicity, monopoly profits are positive and duopoly profits are negative. That is, a firm will never enter the market if it believes that the competitor will enter with sufficiently high probability. For instance, "never enter in $A^M(x, \beta)$" is rational behavior ($h((0,0), x, e) = 1$ for $e \in A^M(x, \beta)$) for both players. However, such behavior cannot be supported by any consistent beliefs. Indeed, if firm 1 believes that firm 2 will not enter, then the former would prefer to enter the market since monopoly profits are positive.

Example 3.1 (Selection mechanism and level-2 rationality). Assume that x, e, and β are such that $e \in A^M(x, \beta)$. Let $\sigma_i \in [0,1]$ be some mixed strategy played by firm i (σ_i is the probability that firm i plays $y_i = 0$). Then in the region of multiplicity, the set of level-2 rational equilibria is $\{(\sigma_1, \sigma_2) \mid \sigma_i \in [0,1],\ i = 1, 2\}$. Let $\rho(\cdot|x, e) : [0,1]^2 \to [0,1]$ be some fixed selection mechanism. That is, ρ is some probability measure over the set of level-2 rational equilibria, $[0,1]^2$. Then the distribution of play is determined by

$$
\begin{aligned}
h((0,0),x,e) &= \int_{[0,1]^2} \sigma_1\sigma_2 d\rho(\sigma_1,\sigma_2|x,e);\\
h((1,0),x,e) &= \int_{[0,1]^2} (1-\sigma_1)\sigma_2 d\rho(\sigma_1,\sigma_2|x,e);\\
h((0,1),x,e) &= \int_{[0,1]^2} \sigma_1(1-\sigma_2) d\rho(\sigma_1,\sigma_2|x,e);\\
h((1,1),x,e) &= \int_{[0,1]^2} (1-\sigma_1)(1-\sigma_2) d\rho(\sigma_1,\sigma_2|x,e).
\end{aligned}
$$

Note that I am not assuming that firms are playing the same equilibrium given the realization of $\mathbf{x}$ and $\mathbf{e}$. Moreover, the selection mechanism is allowed to depend on $\mathbf{x}$ and $\mathbf{e}$ nonparametrically.

Since under level-2 rationality the only region in which h is unknown is $A^M(x,\beta)$, I restrict the domain of h to be $X \times E'$, where

$$
E' = \left\{ e \in E \mid \inf_{x\in X,\beta\in B} (\bar{x}_i^{\mathsf{T}}\bar{\beta}_i - \tilde{x}_i^{\mathsf{T}}\tilde{\beta}_i) \le e_i \le \sup_{x\in X,\beta\in B} \bar{x}_i^{\mathsf{T}}\bar{\beta}_i,\ i \in I \right\}.
$$

Note that $A^M(x,\beta)$ is a strict subset of E' for all x and β.

Under the level-2 rationality assumption the conditional density implied by parameter θ is:

$$
p(y|x;\theta) = \int_{A(y,x,\beta)} f_{e|x}(e|x;\beta)de + \int_{A^M(x,\beta)} h(y,x,e) f_{e|x}(e|x;\beta)de. \tag{5}
$$

3.2. NE constraints

Every NE is level-2 rational. Thus, the distribution of play implied by the NE assumption can be different from the distribution of play implied by the level-2 rationality assumption *only* when $e \in A^M(x,\beta)$. Not surprisingly, the set $A^M(x,\beta)$ is also a region of multiplicity of NE. There are three NE: two in pure strategies and one in mixed strategies. Two pure NE correspond to the monopoly outcomes, $y=(1,0)$ and $y=(0,1)$. Denote the distributions over the outcomes implied by these pure strategy NE by $\alpha_{(1,0)} = (0,0,1,0)^{\mathsf{T}}$ and $\alpha_{(0,1)} = (0,0,0,1)^{\mathsf{T}}$, respectively. Similarly, the mixed-strategy NE constitutes $\alpha : X \times E \times B \to \mathbb{R}^4$ such that

$$
\alpha_1(x,e,\beta) = \left(1 - \frac{\bar{x}_2^{\mathsf{T}}\bar{\beta}_2 - e_2}{\tilde{x}_2^{\mathsf{T}}\tilde{\beta}_2}\right)\left(1 - \frac{\bar{x}_1^{\mathsf{T}}\bar{\beta}_1 - e_1}{\tilde{x}_1^{\mathsf{T}}\tilde{\beta}_1}\right),
$$

$$\alpha_2(x,e,\beta) = \frac{\bar{x}_2^\intercal \bar{\beta}_2 - e_2}{\tilde{x}_2^\intercal \tilde{\beta}_2} \frac{\bar{x}_1^\intercal \bar{\beta}_1 - e_1}{\tilde{x}_1^\intercal \tilde{\beta}_1},$$
$$\alpha_3(x,e,\beta) = \frac{\bar{x}_2^\intercal \bar{\beta}_2 - e_2}{\tilde{x}_2^\intercal \tilde{\beta}_2} \left(1 - \frac{\bar{x}_1^\intercal \bar{\beta}_1 - e_1}{\tilde{x}_1^\intercal \tilde{\beta}_1}\right),$$
$$\alpha_4(x,e,\beta) = \left(1 - \frac{\bar{x}_2^\intercal \bar{\beta}_2 - e_2}{\tilde{x}_2^\intercal \tilde{\beta}_2}\right) \frac{\bar{x}_1^\intercal \bar{\beta}_1 - e_1}{\tilde{x}_1^\intercal \tilde{\beta}_1}.$$

Note that in the multiplicity region, α is a proper distribution over outcomes. Denote vector $(h(y,x,e))_{y\in Y}$ by $h(x,e)$. Since I do not want to impose any parametric restrictions on how players randomize among different NE, $h(x,e)$ can be equal to any convex combination of $\alpha(x,e,\beta)$, $\alpha_{(1,0)}$ and $\alpha_{(0,1)}$.

Definition 5 A pair $\theta_{\text{NE}} = (\beta_{\text{NE}}, h_{\text{NE}}) \in \Theta$ is consistent with the NE assumption if it is consistent with level-2 rationality, and for every x and e such that $e \in A^M(x, \beta_{\text{NE}})$, the following holds

$$h_{\text{NE}}(x,e) \in \text{co}\left(\alpha_{(1,0)}, \alpha_{(0,1)}, \alpha(x,e,\beta_{\text{NE}})\right), \tag{6}$$

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

Example 3.2 (Selection mechanism and NE). Assume that x, e, and β are such that $e \in A^M(x,\beta)$. Then there are three distributions over the outcomes that correspond to three NE of the game: $\alpha_{(1,0)}$, $\alpha_{(0,1)}$ and $\alpha(x,e,\beta)$. Let $\boldsymbol{\rho}_{(1,0)}$ and $\boldsymbol{\rho}_{(0,1)}$ be some *random* probabilities that equilibrium $y = (1,0)$ or equilibrium $y = (0,1)$ is selected. Then the distribution of play is determined by

$$\begin{aligned}
h((0,0),x,e) &= \left(1 - \mathbb{E}\left[\boldsymbol{\rho}_{(1,0)}|x,e\right] - \mathbb{E}\left[\boldsymbol{\rho}_{(0,1)}|x,e\right]\right)\alpha_1(x,e,\beta);\\
h((1,0),x,e) &= \mathbb{E}\left[\boldsymbol{\rho}_{(1,0)}|x,e\right] + \left(1 - \mathbb{E}\left[\boldsymbol{\rho}_{(1,0)}|x,e\right] - \mathbb{E}\left[\boldsymbol{\rho}_{(0,1)}|x,e\right]\right)\alpha_2(x,e,\beta);\\
h((0,1),x,e) &= \mathbb{E}\left[\boldsymbol{\rho}_{(0,1)}|x,e\right] + \left(1 - \mathbb{E}\left[\boldsymbol{\rho}_{(1,0)}|x,e\right] - \mathbb{E}\left[\boldsymbol{\rho}_{(0,1)}|x,e\right]\right)\alpha_3(x,e,\beta);\\
h((1,1),x,e) &= \left(1 - \mathbb{E}\left[\boldsymbol{\rho}_{(1,0)}|x,e\right] - \mathbb{E}\left[\boldsymbol{\rho}_{(0,1)}|x,e\right]\right)\alpha_4(x,e,\beta).
\end{aligned}$$

As one can see, any selection mechanism corresponds to a distribution of play that is a convex combination of the NE distributions over the outcomes. Because of randomness in the selection mechanism, firms in two different markets may end up playing different NE even if they have the same realizations of $\mathbf{x}$ and $\mathbf{e}$.

Let NE $\subseteq \Theta$ be the set of all θ that are consistent with NE. The NE constraints (equation (6)) have to be satisfied for every x and e. Thus, the NE constraints are

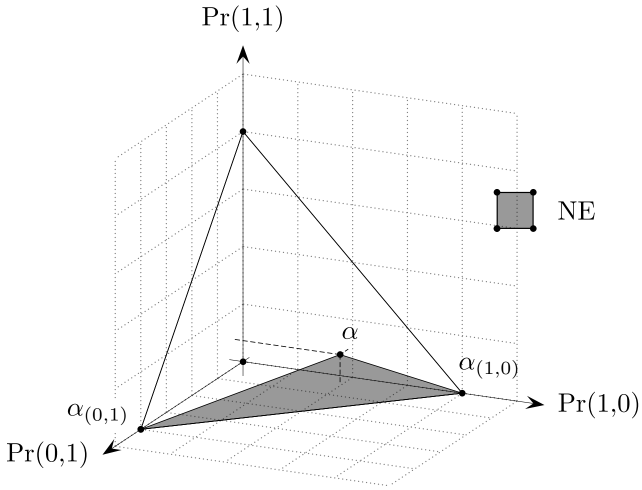


Figure 2 – Predictions for NE in the multiplicity region for fixed x, e and β. The point α corresponds to the mixed NE. The gray region is characterized by an intersection of three half spaces and one hyperplane.

infinite-dimensional. However, the next proposition establishes an equivalent characterization of NE via finite set of constraints.

Proposition 3.1 *Under assumptions 1 - 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 *is 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 ■

I turn the set of infinite-dimensional constraints into the set of finite-dimensional constraints using the set of indicator functions presented in Andrews and Shi (2013). I can do so because of a simple observation that for any function $g : X \times E \to \mathbb{R}$, $g(x, e) \geq 0$ for all $x \in X, e \in E$ if and only if $\mathbb{E}[\, g(\mathbf{x}, \mathbf{e}) | \mathbf{x}, \mathbf{e}\,] \geq 0$ almost surely with respect to any (user-specified) probability measure supported on $X \times E$. Thus, in contrast to Andrews and Shi (2013), the new NE constraints are deterministic and do not need to be estimated. The only role of Proposition 3.1 is to establish an equivalent finite-dimensional characterization of the original infinite-dimensional constraints. Since the constraints in Proposition 3.1 are smooth in β and affine in h, the new constraints are easier to analyze than the original infinite-dimensional ones.

The idea behind Proposition 3.1 is as follows. The convex hull of NE can be characterized as an intersection of one hyperplane, $m_0(\theta) = 0$, and three half spaces, $m_j(\theta) \geq 0$, $j = 1, 2, 3$ (see figure 2). If none of the inequality constraints is binding, then the firms play every NE with a positive probability; if m_1 is binding, then agents never play $\alpha_{(1,0)}$; if m_2 is binding, then agents never play $\alpha_{(0,1)}$; if m_3 is binding, then agents never play the mixed-strategy NE. Note that at most two inequality constraints can be binding simultaneously.

4. Parameter space and identified set

Since under level-2 rationality h_0 is unknown only in the multiplicity region (which is a compact set for every x and β), 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.

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 $|\lambda| = \sum_{i=1}^{d_z} \lambda_i$, $D^\lambda g(m, z) = \partial^{|\lambda|} g(m, z)/\partial z_1^{\lambda_1} \ldots \partial z_{d_z}^{\lambda_{d_z}}$. Then, for η and $\eta_0 > 0$, I define the following modification of the norms which are used in Santos (2012):

$$\|g\|_s = \sqrt{\max_{m \in M} \sum_{|\lambda| \leq \eta + \eta_0} \int_Z [D^\lambda g(m, z)]^2 dz},$$
$$\|g\|_c = \max_{m \in M} \max_{|\lambda| \leq \eta} \sup_{z \in Z} \left| D^\lambda g(m, z) \right|.$$

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

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

where $\min\{\eta, \eta_0\} > d_x/2 + 1$, and K is a finite constant.

Assumption 3 imposes restrictions on what players can do in the multiplicity region. Since every NE outcome distribution (including $\alpha(\cdot, \cdot, \beta)$) is smooth in x and e, the condition that h is at least $(d_x/2 + 1)$-times differentiable is equivalent to assuming the same level of smoothness of the conditional expectation of the selection mechanism (see example 3.2). In the following example Assumption 3 is not satisfied.

Example 4.1 Suppose that in the multiplicity region firms choose to play $y = (1, 0)$ if the monopoly payoff of firm 1 is greater ot equal to the monopoly payoff of firm 2, and firms play $y = (0, 1)$ otherwise. That is,

$$\begin{aligned} h((0,0), x, e) &= 0, \\ h((1,0), x, e) &= \mathbb{1}(\bar{x}_1^\intercal \bar{\beta}_1 - e_1 \geq \bar{x}_2^\intercal \bar{\beta}_2 - e_2), \\ h((0,1), x, e) &= \mathbb{1}(\bar{x}_1^\intercal \bar{\beta}_1 - e_1 < \bar{x}_2^\intercal \bar{\beta}_2 - e_2), \\ h((1,0), x, e) &= 0, \end{aligned}$$

for all x and e such that $e \in A^M(x, \beta)$.

Note that Θ is compact under $\|\theta\|_c = \|\beta\|_e + \|h\|_c$ (see Lemma A.2. in Santos (2012)).

Definition 6 The identified set is

$$\Theta_0 = \{\theta \in \Theta \mid p(\mathbf{y}|\mathbf{x}; \theta) = p_0(\mathbf{y}|\mathbf{x}) \text{ a.s.}\}, \tag{7}$$

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

Usually the identified set is not a singleton.[10] 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:

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

[10]Parametric restrictions on h in the multiplicity region are necessary for point identification of β, when X is a bounded set and there are exclusion restrictions; see Proposition 5.1 in Kashaev and Salcedo (2016).

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 wright $\chi(\theta_1, \theta_2)$ 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

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

Assumption 4 (Rank Condition)

$$\inf_{\theta_0 \in \Theta_0 \bigcap \mathrm{NE}} \lambda_{\min}(\theta_0) > 0. \tag{8}$$

Assumption 4 is a standard regularity condition for parametric likelihood models. Indeed, when h_0 is known, there is only a finite-dimensional parameter β, and $\Lambda(\theta_0)$ is the Fisher information matrix that corresponds to a parametric likelihood model. Thus, Assumption 4 requires that β_0 is locally pointidentified and admits a $\sqrt{n}$-consistent estimator. I do not require neither local nor global identification of β_0. It is possible that for every $(\beta_0, h_0) \in \Theta_0 \bigcap \mathrm{NE}$ there exists $(\beta', h') \in \Theta_0 \bigcap \mathrm{NE}$ such that β' is arbitrary close to β_0. Restrictions on B and existence of the player specific covariates with large but bounded support imply local poitidentification of β_0 given h_0 (see Proposition C.1 in the appendix).[11] To see the necessity of restrictions on B consider the following example.

Example 4.2 Assumption 4 is violated if and only if

$$\exists \theta_0 \in \Theta_0 \bigcap \mathrm{NE},\; \exists \xi \in \mathbb{R}^{d_\beta} \setminus \{0\} \;:\; \forall y \in Y \setminus \{(0,1)\},\; \forall x \in X',\; \xi^\mathsf{T} \partial_\beta p(y|x, \theta_0) = 0.$$

Note that we can focus on $Y \setminus \{(0,1)\}$ since $p((0,1)|x, \theta_0) = 1 - \sum_{y \in Y \setminus \{(0,1)\}} p(y|x, \theta_0)$. Hence, in order to satisfy Assumption 4 one would require to have enough variation in x. Indeed, suppose it is known that

$$u_i(y, x, e, \beta_0) = (\bar{\beta}_{0i} - \tilde{\beta}_{0i} y_{-i} - e_i) y_i,$$

and that h_0 and $f_{e|x}$ do not depend on x. Then there always exists nonzero ξ such that

[11] If there are player specific covariates with full support and $\mathbf{e}$ is independent of $\mathbf{x}$, bivariate normal with *unknown* correlation, then the payoff distribution (including the correlation parameter) is gobally pointidentified under level-2 rationality (see Kashaev and Salcedo (2016)).

$\xi^{\mathsf{T}}\partial_\beta p(y|\theta_0) = 0$ for all $y \in Y \setminus \{(0,1)\}$, since $d_\beta \geq 4$ (four utility function parameter and extra parameters entering $f_{e|x}$) and we at most have three linearly independent equations corresponding to different $y \in Y \setminus \{(0,1)\}$.

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

$$\|\beta\|_{\theta_0} = \sqrt{\beta^{\mathsf{T}}\Lambda(\theta_0)\beta}$$

in $\mathbb{R}^{d_\beta}$ for every $\theta_0 \in \Theta_0 \bigcap \mathrm{NE}$. This norm is the finite-dimensional part of the Fisher norm and depends on the point in the identified set where the variance of the score function is evaluated. It plays a key role in the linear approximation of the NE constraints and the density function.

5. Sieve maximum likelihood setting

The sieve space for h takes the form

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

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 $\Theta_{k(n)} = B \times \mathcal{H}_{k(n)-d_\beta}$.

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 $p_0 f_x$.

Assumption 6 is standard in the literature. However, it might be violated if there is a correlation in observed outcomes in different markets caused by an unobserved correlation in the selection mechanism.[12]

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 and γ_n be the sieve approximation rate. That is, $L_n(\theta) = \sum_{i=1}^n \log p(\mathbf{y}_i|\mathbf{x}_i;\theta)$, and for every $h \in \mathcal{H}$ there exists $h_k \in \mathcal{H}_{k(n)-d_\beta}$ such that for some $c_1 \in \mathbb{R}$

$$\|h_k - h\|_\infty \leq c_1 \gamma_n. \tag{9}$$

Theorem 5.1 *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 1–3, 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)

$$\sup_{\tilde{\theta}_n \in \tilde{\Theta}_n} \inf_{\theta_0 \in \Theta_0 \bigcap \mathrm{NE}} \left\|\tilde{\beta}_n - \beta_0\right\|_{\theta_0} = o_p(n^{-1/4}). \tag{10}$$

Proof. See Appendix. ■

Results 1 and 2 are applications of Theorem 3.1 of CTT. The last result is new to the literature and is essential for testing the NE constraints. It permits linearization

[12]See Epstein et al. (2015).

of the constraints in β, which, together with affinity in h, provides the asymptotic distribution of the test statistics introduced in the next section.

6. 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 hypothesis 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 the definition of 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 $\alpha_{(1,0)}$, $\alpha_{(0,1)}$, 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 NE, $\alpha_{(0,1)}$, and $\alpha_{(1,0)}$ is always played in the multiplicity region.

Define the following sieve-LR 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 null hypothesis, the above statistics have a tight limit.

Theorem 6.1 *Under the assumptions of Theorem 5.1,*

(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 6.1, 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 $\alpha_{(1,0)}$ 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. In general, the asymptotic null distribution of $T_{n,j}$ is the supremum of a chi-squared process. In the point-identified case, similar to $T_{n,i,j}$, it reduces to the chi-squared 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 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 $T_{n,j}$, $j = 1,2,3$. If $T_{n,i} \leq c_{n,i}(1-\alpha)$ for some $j = 1,2,3$, then accept the null hypothesis. Otherwise proceed to Step 3.

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

The procedure is conservative. However, if the null hypothesis of Step 1 is rejected, then the players are likely to always mix between different NE; if the null hypothesis of Step 2 is rejected, then the players are likely to always mix between all NE.[13] Thus, as an intermediate step one can learn what agents do in the region of multiplicity.

7. Multiplier bootstrap

Once the asymptotic null distribution of each statistic is derived, I show that the multiplier bootstrap with i.i.d weights approximates quantiles of a distribution that dominates the asymptotic null distribution of the statistic. 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, $T^{w,r}_{n,0}$; *(iii)* Use

$$c_{n,0}(1-\alpha) = \inf\left\{\tau : \frac{1}{R_n}\sum_{r=1}^{R_n} \mathbb{1}(T^{w,r}_{n,0} \leq \tau) \geq 1-\alpha\right\} \tag{11}$$

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 $\{(\mathbf{y}_i, \mathbf{x}_i)\}_{i=1}^n$.

Recall that $\hat{\theta}_n$ is the 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

$$L^w_n(\theta) = \sum_{i=1}^n \mathbf{w}_i \log\left(p(\mathbf{y}_i|\mathbf{x}_i,\theta)\right),$$

[13] The order of steps is not important since the null hypotheses in Theorem 6.1 are mutually exclusive.

$$T^w_{n,0} = 2\left[\sup_{\theta\in\Theta_{k(n)}} L^w_n(\theta) - \sup_{\theta\in\Theta_{k(n)}\cap\mathrm{NE}(\hat{\theta}_n)} L^w_n(\theta)\right],$$

$$T^w_{n,j} = 2\left[\sup_{\theta\in\Theta_{k(n)}} L^w_n(\theta) - \sup_{\theta\in\Theta_{k(n)}\cap\mathrm{NE}_i(\hat{\theta}_n)} L^w_n(\theta)\right], j = 1, 2, 3,$$

$$T^w_{n,i,j} = 2\left[\sup_{\theta\in\Theta_{k(n)}} L^w_n(\theta) - \sup_{\theta\in\Theta_{k(n)}\cap\mathrm{NE}_{i,j}(\hat{\theta}_n)} L^w_n(\theta)\right], (i, j) \in \{(1, 2), (1, 3), (2, 3)\}.$$

In contrast to CTT and Tao (2014), my bootstrap statistics are not centered around points in the restricted identified set ($\Theta_0 \cap \mathrm{NE}$) since one cannot guarantee that the NE constraints will be even approximately satisfied at $\hat{\theta}_n$. CTT assume that the constraint can be linearized around points in the restricted identified set.[14] Tao (2014) assumes existence of a particular reparametrization of the constraints, and then requires that these reparametrized constraints can be linearized around points in the restricted identified set. Such a linearization allows them to show that the bootstrap distribution approximates the limiting null distribution of their test statistics. In my setting, the constraints can be linearized around point in the *unrestricted* identified set (Θ_0). As a result, my bootstrap distribution will dominate the limiting hull distribution. Thus, the bootstrap critical values can be conservative.

In order to be able to linearize the constraints around points in Θ_0 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)

$$\inf_{\theta_0\in\Theta_0} \lambda_{\min}(\theta_0) > 0. \tag{12}$$

Theorem 7.1 *Under the assumptions of Theorem 6.1, assumption 7–8,*

(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 (11)

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

[14] In the case with linear constraints (e.g., $m_0(\theta) = \theta$) this assumption is trivially satisfied.

(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.*

8. Extensions

The proposed procedure can be easily extended to other binary games and to binary games with more than two players.

8.1. Coordination game

The coordination game in this setting is different from the entry game in one respect: the competition effect $\tilde{x}_i^\intercal \tilde{\beta}_i < 0$. Thus, the level-2 rationality predictions are slightly different. In the multiplicity region $A^M(x,\beta)$, the level-2 rationality assumption does

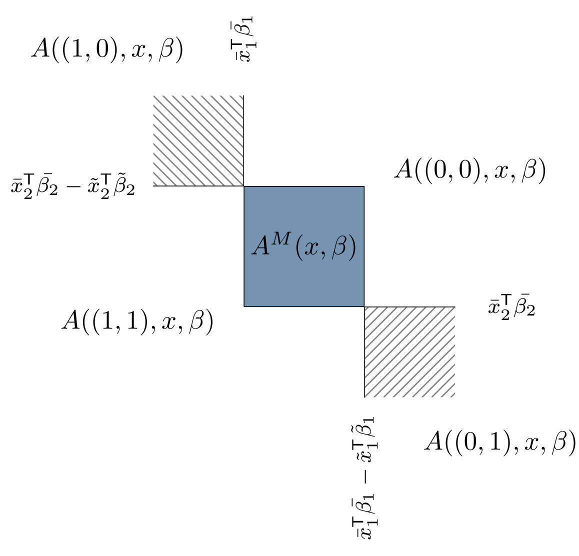


Figure 3 – Possible outcomes of the coordination game under the level-2 rationality assumption.

not impose any restrictions on h. However, under the NE assumption, there are three

Nash equilibria in the multiplicity region: two pure strategy equilibria ($(1,1)$ and $(0,0)$) and the mixed-strategy equilibrium. Hence, one can build the likelihood function and the NE constraints following the same steps as in the case of the entry game.

8.2. Games with more than two players

The main intuition of a two-player entry game can be applied to games with more than two players. However, there are two major differences. Given x and β, *(i)* there are more than one multiplicity region; and *(ii)* some of the multiplicity regions are not bounded sets. Indeed, assume that the payoff of player i for an outcome y is given by

$$u_i(y,x,e,\beta) = (\bar{x}_i^{\intercal}\bar{\beta}_i - \tilde{x}_i^{\intercal}\tilde{\beta}_i \sum_{j\neq i} y_j - e_i)y_i. \tag{13}$$

When there are three players and $e_1 > \bar{x}_1^{\intercal}\bar{\beta}_1$, player 1 will never enter the market. Knowing that firm 1 is not entering, the other firms will be playing a two-player entry game. Hence, the unbounded set

$$\{e \in \mathbb{R}^3 : e_1 > \bar{x}_1^{\intercal}\bar{\beta}_1, \bar{x}_i^{\intercal}\bar{\beta}_i - \tilde{x}_i^{\intercal}\tilde{\beta}_i \leq e_i \leq \bar{x}_i^{\intercal}\bar{\beta}_i, i = 2,3\} \tag{14}$$

is a region of multiplicity of NE.

Similarly,

$$\{e \in \mathbb{R}^3 : e_2 > \bar{x}_2^{\intercal}\bar{\beta}_2, \bar{x}_i^{\intercal}\bar{\beta}_i - \tilde{x}_i^{\intercal}\tilde{\beta}_i \leq e_i \leq \bar{x}_i^{\intercal}\bar{\beta}_i, i = 1,3\} \tag{15}$$

is another unbounded multiplicity region.

Assumption 9 For almost every $x \in X$ and every $\beta \in B$, if for player $i \in I$, $y_i = 1$ or $y_i = 0$ is a strictly dominant strategy, then $h(y,x,e)$ does not depend on e_i for all $y \in Y$.

Assumption 9 guarantees that, in the multiplicity regions, h depends only on variables that have a compact support. Next, one can show that under the NE assumption, E can be partitioned into (i) the sets in which, NE and level-2 rationality give identical point predictions, and (ii) the sets in which level-2 rationality does not impose any restrictions on h, but the NE assumption does. Formally:

$$p(y|x;\theta) = \int_{A(y,x,\beta)} f_{e|x}(e|x;\beta)de + \sum_{k=1}^{K} \int_{A_k^M(x,\beta)} h_k(y,x,e) f_{e|x}(e|x;\beta)de, \qquad (16)$$

where K is a number of the multiplicity regions. Every multiplicity region, $A_k^M(x,\beta)$, imposes a set of equality/inequality constraints on h_k. For instance, in the multiplicity region defined by equation (14), $(h_k((1,y_{-1}),x,e))_{y_{-1}} = 0$ and $(h_k((0,y_{-1}),x,e))_{y_{-1}} \in [0,1]^4$ has to satisfy the constraints defined in Proposition 3.1. As a result, a new parameter vector is $\theta = (\beta, h_1, \ldots, h_K)$; the set of NE constraints is $\{m_{kj}(\theta) \geq (=)0, j = 0,1,2,\ldots\}_{k=1}^K$; and the sequence of the test statistics is defined by all possible cases with binding inequality constraints. Unfortunately, in games with a large number of players the number of the test statistics can be too large. However, one can increase the computational tractability of the problem by testing for NE in pure strategies only.

8.3. Confidence sets for payoff parameters

In this section I show how one can couple my testing procedure with the LR test of CTT in order to construct confidence sets for the payoff parameters. Suppose that one wants to test the null hypothesis that $\beta = r_0$ under the level-2 rationality assumption. My set up allows one to use the test statistic suggested by CTT:

$$\mathrm{LR}^{\mathrm{l2r}}(r_0) = 2\left[\sup_{\theta\in\Theta_{k(n)}} L_n(\theta) - \sup_{\theta\in\Theta_{k(n)}\cap\{\beta=r_0\}} L_n(\theta)\right]. \qquad (17)$$

Note that $\sup_{\theta\in\Theta_{k(n)}} L_n(\theta)$ has already been computed in Step 1 of my procedure. Thus, one just needs to recompute $\sup_{\theta\in\Theta_{k(n)}\cap\{\beta=r_0\}} L_n(\theta)$. The later is a concave optimization problem if one uses linear sieves. The same computational simplification can be applied to the computation of the bootstrap statistics of CTT.

Similarly, if one wants to test the null hypothesis that $\beta = r_0$ under the NE assumption, she can use

$$\mathrm{LR}^{\mathrm{NE}}(r_0) = 2\left[\sup_{\theta\in\Theta_{k(n)}\cap\mathrm{NE}} L_n(\theta) - \sup_{\theta\in\Theta_{k(n)}\cap\mathrm{NE}\cap\{\beta=r_0\}} L_n(\theta)\right]. \qquad (18)$$

Note that the first supremum in (18) has been computed in Step 3 of my procedure, and computation of the second supremum in (18) is a concave maximization problem with linear constrains on h.

The confidence sets for β can be obtained by inversion of $\mathrm{LR}^{\mathrm{l2r}}$ and $\mathrm{LR}^{\mathrm{NE}}$.

9. Monte Carlo experiments, and empirical application

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

9.1. The Monte Carlo experiments

I provide results for the finite-sample performance of my testing procedure (sLR).[15] I also test the NE assumption by using the BMM approach, coupled with that of Andrews and Shi (2013)(BMM+AS).[16] I consider a two-player entry game of complete information without covariates. 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 reasons.[17] First, I want to evaluate the effect of the sieve approximation of the distribution of play on the finite-sample performance of my procedure. Second, in my Monte Carlo experiments the testing procedure of Andrews and Shi (2013) numerically coincides with the testing procedures of Andrews and Soares (2010) and Bugni et al. (2015) (if one uses the same generalized moment selection procedure). As a result, Andrews and Shi (2013)'s "by product" testing procedure is numerically equivalent to the model specification test of Bugni et al. (2015).

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

[15] In the last section of the appendix I discuss computational aspects of my procedure and provide a pseudo-algorithm to compute the test statistics.

[16] I directly follow the recommendations of Andrews and Shi (2013) in computing their test statistic and critical values.

[17] I also conducted Monte Carlo simulations assuming that some of the elements of β_0 are unknown under the null hypothesis. The results of these experiments are similar to the results obtained in the case when β_0 is known.

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

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

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

$$\max_{p\in[0,1]} \Big[p u_1((1,0), e) \Big]^{\phi(e)} \Big[(1-p) u_2((0,1), e) \Big]^{1-\phi(e)}.$$

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

Under the alternative in which the NE assumption is false, I consider five 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.[18] 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.[19] 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.

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

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

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 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, #bootstrap = 500.

Table 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

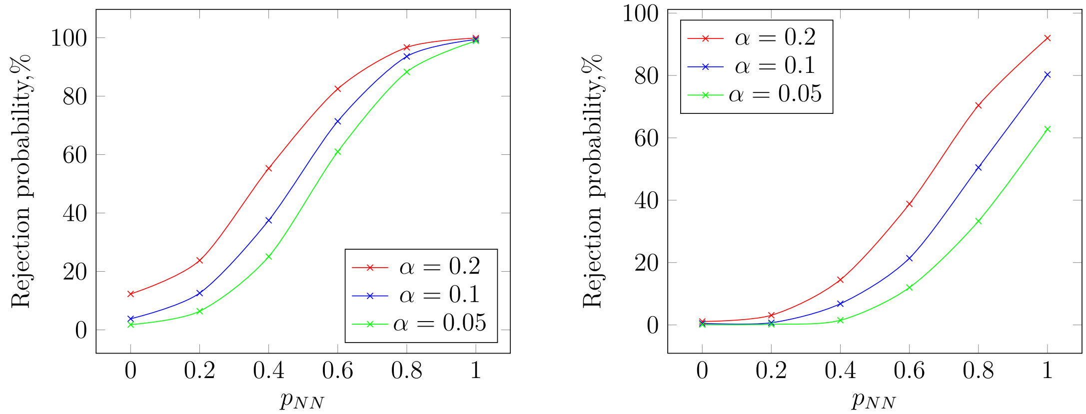
Notes: α is a significance level; #MC = 1000, #bootstrap = 500.

Both approaches are conservative. However, the proposed sieve LR procedure is substantially less conservative. It is worth noting that the procedure based on BMM with AS (i) is not designed to test the NE assumption, (ii) is uniform, and (iii) does not assume smoothness of the distribution of play.

The power results are presented in figures 4 and 5. The power of both procedures improves with an increase in sample size. The "sLR" procedure has better power than the "BMM+AS" procedure.

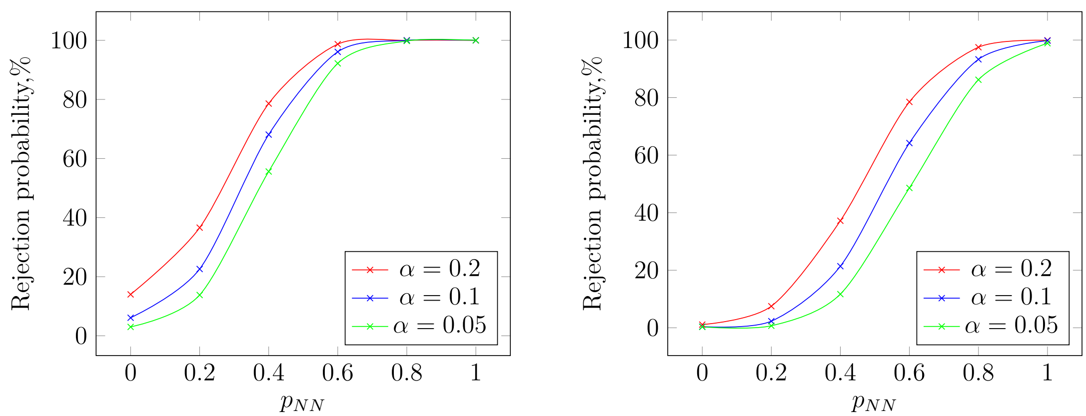
9.2. An empirical application

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



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

Figure 4 – Power curves, $n = 500$



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

Figure 5 – Power curves, $n = 1000$

patterns of small rural grocery stores in the US between 1998 and 2002. I now consider a complete information version of that model, assuming that the unobserved payoff shifters are perfectly correlated. I also tested the model allowing for unknown correlation between the unobserved payoff shifters. The value of the correlation parameter that maximizes objective function under both level-2 rationality and the NE assumption is equal to 1. Thus, I encounter the parameter on the boundary problem. To overcome this problem I decided to assume the correlation parameter to be equal to 1. Perfectly correlated unobserved payoff shifters imply that there is only market specific heterogeneity.

For each market, the data I use contains information on whether each firm was active in 2002, the population, the distance from a supercenter (Walmart), and whether each firm was active in 1998. In total there are 4803 observations.[20] The 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

$$\begin{aligned}\mu(x, \beta) =& \beta_0 + \beta_1 \mathbb{1}(pop > 3k) + \beta_2 \mathbb{1}(pop > 6k) + \\ & \beta_3 \mathbb{1}(supercenter < 20mi) + \beta_4 \mathbb{1}(iInactive1998), \\ \delta(x, \beta) =& \beta_5 + \beta_6 \mathbb{1}(pop > 3k) + \beta_7 \mathbb{1}(pop > 6k) + \beta_8 \mathbb{1}(supercenter < 20mi),\end{aligned}$$

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

I use fifth order degree polynomials to approximate the distribution of play. To 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. I use 200 bootstrap replications to approximate the critical values. The NE assumption is rejected at the 5 percent significance level.

Table 3 – Sample log-likelihood

	level-2 rationality	All	no (1,0)	no (0,1)	no mixed	mixed only
Log-likelihood	-2974	-3099	-3443	-3380	-10250	-3580

Notes: "level-2 rationality" - the value of log-likelihood under the level-2 rationality 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.

10. Conclusion

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

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

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 noting that the asymptotic results presented in the paper hold under a fixed data generating process. Since the asymptotic null 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 future work.

References

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

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., and Ryan, S. P. (2010). Identification and estimation of a discrete game of complete information. *Econometrica*, 78(5):1529–1568.

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 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.

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

Camerer, C. (2003). *Behavioral Game Theory: experiments in strategic interaction*. Princeton University Press.

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*.

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

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.

Epstein, L. G., Kaido, H., and Seo, K. (2015). Robust confidence regions for incomplete models. *Econometrica, forthcoming*.

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

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.

Henry, M. and Mourifie, I. (2012). Nonparametric sharp bounds for payoff functions in 2x2 games. Mimeo.

Kashaev, N. and Salcedo, B. (2016). Identification of solution concepts for discrete games. Working paper.

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.

Liu, X. and Shao, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *Annals of Statistics*, pages 807–832.

Murphy, S. A. and Van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–465.

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.

Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence*. Springer.

A. Notation and definitions

$$
\begin{aligned}
[d] &= \max\{n \in \mathbb{N}, d \geq n\}, \\
\|\beta\|_e &= \sqrt{\beta^\mathsf{T}\beta}, \\
\|h\|_s^2 &= \max_{y\in Y} \sum_{|\lambda|\leq \eta+\eta_0} \int [D^\lambda h(y,x,e)]^2 dxde, \\
\|h\|_c &= \max_{y\in Y} \max_{|\lambda|\leq \eta} \sup_{x,e} \left| D^\lambda h(y,x,e) \right|, \\
\|\theta\|_s &= \|\beta\|_e + \|h\|_s\,, \\
\|\theta\|_c &= \|\beta\|_e + \|h\|_c\,, \\
\Theta_0 &= \text{the identified set}, \\
p(\theta) &= p(\cdot|\cdot,\theta), \\
p_0 &= p(\theta_0), \quad \theta_0 \in \Theta_0, \\
P_0 &= p_0 f_x, \\
L^2(P_0) &= \left\{ g : Y \times X \to \mathbb{R} : \int g^2 dP_0 < \infty \right\}, \\
\langle g_1, g_2 \rangle_{L^2(P_0)} &= \int g_1 g_2 dP_0, \\
\|g\|_{L^2(P_0)} &= \sqrt{\langle g, g\rangle_{L^2(P_0)}}, \\
\chi^2(P_1,P_2) &= \int \left(\frac{P_1}{P_2} - 1 \right)^2 P_2 d\mu, \text{ where } \mu \text{ is a dominating } \sigma\text{-finite positive measure}, \\
\chi^2(\theta,\theta_0) &= \chi^2(p(\theta),p_0) = \int \left(\frac{p(\theta)}{p_0} - 1 \right)^2 P_0 d\mu = \left\| \frac{p(\theta)}{p_0} - 1 \right\|^2_{L^2(P_0)}, \\
H^2(P_1,P_2) &= \int (\sqrt{P_1} - \sqrt{P_2})^2 d\mu, \\
\mathcal{G} &= \left\{ g \in L^2(P_0) : \|g\|_c < C \right\}, \text{ for some known } C < \infty, \\
\bar{\mathcal{G}} &= \left\{ g \in \mathcal{G} : \int g dP_0 = 0, \|g\|_{L^2(P_0)} = 1 \right\}, \\
\mu_n\{g\} &= n^{-1} \sum_{i=1}^n g(\mathbf{y}_i, \mathbf{x}_i), \\
\text{NE} &= \{\theta \in \Theta : \theta \text{ satisfies NE constraints}\}, \\
\text{NE}_j &= \Theta_0 \bigcap \{\theta \in \Theta : m_j(\theta) = 0\},
\end{aligned}
$$

B. Level-2 rationality 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 $i \in I$ denote

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

Then under the level-2 rationality assumption

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

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 (6), for $\theta = (\beta, h)$ to be consistent with the NE assumption, the following has to be true uniformly over the multiplicity region:

$$h(x,e) \in \text{co}\Big((0,0,1,0)^\top, (0,0,0,1)^\top, \alpha(x,e,\beta)\Big).$$

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

$$\tilde{m}_0(x,e,\theta) = \alpha((1,1),x,e,\beta)h((0,0),x,e) - \alpha((0,0),x,e,\beta)h((1,1),x,e), \tag{19}$$

$$\tilde{m}_1(x,e,\theta) = -\alpha((1,0),x,e,\beta)h((1,1),x,e) + \alpha((1,1),x,e,\beta)h((1,0),x,e), \tag{20}$$

$$\begin{aligned}\tilde{m}_2(x,e,\theta) =& \frac{\alpha((1,1),x,e,\beta)+\alpha((1,0),x,e,\beta)-1}{\alpha((0,0),x,e,\beta)}h((0,0),x,e) - \\ &- h((1,1),x,e) - h((1,0),x,e) + 1,\end{aligned} \tag{21}$$

$$\tilde{m}_3(x,e,\theta) = h((1,0),x,e) + h((0,1),x,e). \tag{22}$$

Then

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

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 $(e^\top, x^\top)^\top \in \mathbb{R}^{d_e+d_x}$ and side length less than $2\bar{r}$. Let $z = (e^\top, x^\top)^\top$. Then

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

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

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

where $\Phi(\cdot)$ is the standard normal c.d.f., and $\mathbb{E}_\mu [\,\cdot\,]$ is taken with respect to any probability measure μ supported on $X \times E$.

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

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

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

C. Sufficient conditions for local pointidentification of β_0 in Assumption 4

Assumption 4 requires local indentification of β_0 and $\sqrt{n}$-consistensy of the MLE estimator of β_0 when h_0 is known. Below I provide conditions that ensure that if h_0 is known, then β_0 is locally pointidentified.

Proposition C.1 *Let h_0 be a known function such that $(\beta, h_0) \in \Theta_0 \bigcap \text{NE}$ for some $\beta \in B$. Under assumptions 1-3, if (i) $\mathbf{e}|\mathbf{x}$ is bivariate mean-zero normal with variances equal to 1, and unknown correlation $\beta_e \in (-1, 1)$ that does not depend on x; (ii) There are player-specific covariates that enter $v_i(x, \beta) = \bar{x}^\intercal \bar{\beta}$ only; (iii) B is such that the coefficients in front of these player specific covariates are not zero; then $\beta_u = (\bar{\beta}_1^\intercal, \bar{\beta}_2^\intercal, \tilde{\beta}_1^\intercal, \tilde{\beta}_2^\intercal)^\intercal$ is locally pointidentified. If, moreover, (iii) for every $\beta \in B$ such that $\beta_e \in [0, 1)$, with positive probability $A^M(\mathbf{x}, \beta)$ is a subset of one of the following sets*

$$\begin{aligned} &\{e \in \mathbb{R}^2 : e_1 \leq -(1 - \beta_e^2)^3, e_2 \geq 0\} \bigcup \{e \in \mathbb{R}^2 : e_1 \leq 0, e_2 \geq (1 - \beta_e^2)^3\}, \\ &\{e \in \mathbb{R}^2 : e_1 \geq (1 - \beta_e^2)^3, e_2 \leq 0\} \bigcup \{e \in \mathbb{R}^2 : e_1 \geq 0, e_2 \leq -(1 - \beta_e^2)^3\}; \end{aligned}$$

(iv) For every $\beta \in B$ such that $\beta_e \in (-1, 0]$, with positive probability $A^M(\mathbf{x}, \beta)$ is a

subset of one of the following sets

$$\{e \in \mathbb{R}^2 : e_1 \leq -(1-\beta_e^2)^3, e_2 \leq 0\} \bigcup \{e \in \mathbb{R}^2 : e_1 \leq 0, e_2 \leq -(1-\beta_e^2)^3\},$$
$$\{e \in \mathbb{R}^2 : e_1 \geq (1-\beta_e^2)^3, e_2 \geq 0\} \bigcup \{e \in \mathbb{R}^2 : e_1 \geq 0, e_2 \geq (1-\beta_e^2)^3\};$$

then β_e *is also locally pointidentified.*

Proof. We want to show that the set $B(h_0) = \{\beta \in B \mid (\beta, h_0) \in \Theta_0 \bigcap \mathrm{NE}\}$ consists of locally isolated points for every h_0. The proof consists of three steps. In steps 1 and 2, I show that the utility parameter $\beta_u = (\bar{\beta}_1^\intercal, \bar{\beta}_2^\intercal, \tilde{\beta}_1^\intercal, \tilde{\beta}_2^\intercal)^\intercal$ is locally identified. In step 3, I show local identification of the parameters that govern the distribution of unobservables $\mathbf{e}$. Fix some h_0.

Step 1. Take any $\beta \in B(h_0)$. If there exists $X' \subseteq X$ with positive measure such that for every $x \in X'$ there exists $E' \subseteq A^M(\beta, x)$ with positive measure such that $h_0((1,1), x, e) > 0$ for all $e \in E'$, then the utility function parameter $\beta_u = (\bar{\beta}_1^\intercal, \bar{\beta}_2^\intercal, \tilde{\beta}_1^\intercal, \tilde{\beta}_2^\intercal)^\intercal$ is locally identified. Indeed, by way of contradiction assume that for every $\epsilon > 0$ there exists $\beta' \in B(h_0)$ such that $\|\beta - \beta'\| < \epsilon$ and $\beta_u \neq \beta'_u$. Take ϵ small enough such that for every for every $x \in X'$ $E' \bigcap A^M(\beta', x)$ has a positive measure. Hence, for all $x \in X'$, $e \in E' \bigcap A^M(\beta', x)$, and $\beta_0 \in \{\beta, \beta'\}$

$$\frac{h_0((0,0),x,e)}{h_0((1,1),x,e)} = \frac{\alpha((0,0),x,e,\beta_0)}{\alpha((1,1),x,e,\beta_0)} = \frac{(e_1 - v_1(x,\beta_0) + \tilde{v}_1(x,\beta_0))(e_2 - v_2(x,\beta_0) + \tilde{v}_2(x,\beta_0))}{(v_1(x,\beta_0) - e_1)(v_2(x,\beta_0) - e_2)}$$

Note that $\dfrac{h_0((0,0),x,e)}{h_0((1,1),x,e)}$ is identified, since h_0 is assumed to be known. Let

$$\gamma_0(x,e) = \frac{h_0((0,0),x,e)}{h_0((1,1),x,e)}.$$

Then,

$$\gamma_0(x,e) = \left(\frac{\tilde{v}_1(x,\beta_0)}{v_1(x,\beta_0) - e_1} - 1\right)\left(\frac{\tilde{v}_2(x,\beta_0)}{v_2(x,\beta_0) - e_2} - 1\right).$$

Next note that for $i = 1, 2$

$$2\frac{\partial_{e_i}\gamma_0(x,e)}{\partial^2_{e_i^2}\gamma_0(x,e)} = v_i(x,\beta_0) - e_i.$$

Hence, by the rank condition on $\bar{x}_i$ ($v_i(x,\beta_0) = \bar{x}_i^\intercal \bar{\beta}_{0i}$), we conclude that $\bar{\beta}_i = \bar{\beta}'_i$,

$i = 1, 2$. Finally, note that

$$\tilde{v}_i(x, \beta_0) = \frac{(v_i(x, \beta_0) - e_i)^2}{v_i(x, \beta_0) - e_i - \dfrac{\gamma_0(x, e)}{\partial_{e_i}\gamma_0(x, e)}}$$

Hence, by the rank condition on $\tilde{x}_i$ ($\tilde{v}_i(x, \beta_0) = \tilde{x}_i^{\intercal}\tilde{\beta}_{0i}$), we conclude that $\tilde{\beta}_i = \tilde{\beta}'_i$, $i = 1, 2$. The contradiction proves the claim.

Step 2. Note that Step 1 implies that if for $\beta \in B(h_0)$ there exists $X' \subseteq X$ with positive measure such that for every $x \in X'$ there exists $E' \subseteq A^M(\beta, x)$ with positive measure such that $h((1,1), x, e) > 0$ for all $e \in E'$, then the utility function parameter β_u is locally pointidentified. Suppose that $\beta \in B(h_0)$ is such that for all $x \in X$ and for all $e \in A^M(\beta, x)$ $h_0((1,1), x, e) = 0$. That is, the mixed strategy NE is never played in the multiplicity region. Similarly to step 1 assume that for every $\epsilon > 0$ there exists $\beta' \in B(h_0)$ such that $\|\beta - \beta'\| < \epsilon$ and $\beta_u \neq \beta'_u$. Step 1 implies that for all $x \in X$ and for all $e \in A^M(\beta', x)$ $h_0((1,1), x, e) = 0$. If the later is not true, then by step 1 β'_u is locally identified. Hence, we have two points $\beta, \beta' \in B(h_0)$, $\beta \neq \beta'$ such that firms always play PNE. The later is not possible since under our assumptions, under PNE solution concept, β is pointidentified (see Kline (2015b)).

Step 3. We have showed that β_u is locally identified independently of β_e (the correlation parameter). It is left to show that $\partial_{\beta_e} p(\mathbf{y}|\mathbf{x}; (\beta_u^{\intercal}, \beta_e)^{\intercal}, h_0) \neq 0$ with positive probability. Let $\phi(e_1, e_2; \beta_e)$ be the p.d.f. of $\mathbf{e}$. Then

$$\partial_{\beta_e}\phi(e_1, e_2; \beta_e) = \frac{\phi(e_1, e_2; \beta_e)}{\sqrt{1 - \beta_e^2}}\left(\beta_e + \frac{(\beta_e e_2 - e_1)(\beta_e e_1 - e_2)}{(1 - \beta_e^2)^{3/2}}\right).$$

Note that if $b_e \geq 0$, then with positive probability either $\partial_{\beta_e} p((1,0)|\mathbf{x}; (\beta_u^{\intercal}, \beta_e)^{\intercal}, h_0) < 0$ or $\partial_{\beta_e} p((0,1)|\mathbf{x}; (\beta_u^{\intercal}, \beta_e)^{\intercal}, h_0) < 0$. Indeed, take some $x \in X$ such that $A^M(x, \beta)$ is a subset of, say,

$$\{e \in \mathbb{R}^2 : e_1 \leq -(1 - \beta_e^2)^3, e_2 \geq 0\} \bigcup \{e \in \mathbb{R}^2 : e_1 \leq 0, e_2 \geq (1 - \beta_e^2)^3\}.$$

Then, $A((1,0), x, \beta)$ is also a subset of it. Since, for all

$$e \in \{e \in \mathbb{R}^2 : e_1 \leq -(1 - \beta_e^2)^3, e_2 \geq 0\} \bigcup \{e \in \mathbb{R}^2 : e_1 \leq 0, e_2 \geq (1 - \beta_e^2)^3\}$$

$\partial_{\beta_e}\phi(e_1, e_2; \beta_e) < 0$, then

$$\partial_{\beta_e} p((1,0)|\mathbf{x}; (\beta_u^{\intercal}, \beta_e)^{\intercal}, h_0) = \int_{A((1,0),x,\beta)} \partial_{\beta_e}\phi(e_1, e_2; \beta_e) de + \int_{A^M(x,\beta)} h_0((1,0), x, e)\partial_{\beta_e}\phi(e_1, e_2; \beta_e) de$$

$$\leq \int_{A((1,0),x,\beta)\cup A^M(x,b)} \partial_{\beta_e}\phi(e_1, e_2; \beta_e)de < 0.$$

Similarly, one can analyze $\partial_{\beta_e} p((0,1)|\mathbf{x}; (\beta_u^\intercal, \beta_e)^\intercal, h_0)$. Sufficiency of condition (iv) when $\beta_e < 0$ can be proved using the same arguments but for outcomes $(0,0)$ or $(1,1)$. ■

D. Proof of Theorem 5.1

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 (9). 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 $\Theta_0 \bigcap \text{NE}$.

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

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

Proof. Let θ_{0n} be a projection of θ_0 on $\Theta_{k(n)}$. By Theorem 5.1.1-2 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

$$\begin{aligned}
o(n^{-1/4}) = \chi((\beta, h_0), \theta_0) = \left\|\frac{p(\beta, h_0)}{p_0} - 1\right\|_{L^2(P_0)} = \\
&= \left\|\frac{\partial_{\beta^\intercal} p(\theta_0)}{p_0}(\beta - \beta_0) + O\left(\|\beta - \beta_0\|_e^2\right)\right\|_{L^2(P_0)} \geq \\
&\geq |\text{the triangular inequality}| \geq \left|\left\|\frac{\partial_{\beta^\intercal} p(\theta_0)}{p_0}(\beta - \beta_0)\right\|_{L^2(P_0)} - O\left(\|\beta - \beta_0\|_e^2\right)\right| = \\
&= \left|\|\beta - \beta_0\|_{\theta_0} - O\left(\|\beta - \beta_0\|_{\theta_0}^2\right)\right|.
\end{aligned}$$

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

$$\begin{aligned}
&o(n^{-1/4}) = \chi(\theta, \theta_0) = \\
&= \left\|\frac{\Gamma(\beta_0, h - h_0)}{p_0} + \frac{\partial_{\beta^\intercal}\Gamma(\beta_0, h - h_0)(\beta - \beta_0)}{p_0} + \frac{\partial_{\beta^\intercal} p(\theta_0)}{p_0}(\beta - \beta_0) + o_p(n^{-1/2})\right\|_{L^2(P_0)} \geq \\
&\geq \left|\left\|\frac{\Gamma(\beta_0, h - h_0)}{p_0}\right\|_{L^2(P_0)} - o(n^{-1/4})\right|.
\end{aligned}$$

As a result,

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

■

E. 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 $\theta_0 \in \Theta_0 \bigcap \text{NE}$ 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 E.1 *Under assumption of Theorem 5.1*

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

Proof. By definition of $B_n(\theta_0)$, Lemma D.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 E.2 *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 D.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,

$$\begin{aligned} 0 =& m_j(\theta) = m_j(\beta_0, h) - m_j(\beta_0, h_{0n}) + \partial_{\beta^\intercal} m_j(\beta_0, h_{0n})(\beta - \beta_0) + \\ &+ \left(\partial_{\beta^\intercal} m_j(\beta_0, h) - \partial_{\beta^\intercal} m_j(\beta_0, h_{0n})\right)(\beta - \beta_0) + o_p(n^{-1/2}). \end{aligned} \quad (23)$$

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 (23) 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,

$$\sup_{\theta \in B_n^j(\theta_0)} \left| \partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}] + \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}](\beta - \beta_0) \right| = o_p(n^{-1/2}). \quad (24)$$

Note that by the triangular inequality,

$$\begin{aligned}
&\sup_{\theta \in B_n^j(\theta_0)} \left| \partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}] + \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}](\beta - \beta_0) \right| \geq \\
&\sup_{\theta \in B_n^j(\theta_0)} \left| \left| \partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}] \right| - \left| \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}](\beta - \beta_0) \right| \right| \geq \\
&\left| \sup_{\theta \in B_n^j(\theta_0)} \left| \partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}] \right| - \sup_{\theta \in B_n^j(\theta_0)} \left| \partial_{\beta^\intercal} \psi_j(\beta_0)[h - h_{0n}](\beta - \beta_0) \right| \right| = \\
&\sup_{\theta \in B_n^j(\theta_0)} \left| \partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}] \right| \left| 1 - \xi_n \right|, \qquad (25)
\end{aligned}$$

where

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

Since by the Cauchy-Schwarz and the triangular inequalities,

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

and

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

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

The fact that $\xi_n = o_p(1)$ together with (24) and (25) imply that

$$\sup_{\theta \in B_n(\theta_0) \cap \mathrm{NE}_j} \left| \partial_\theta m_j(\theta_{0n})[\theta - \theta_{0n}] \right| = o_p(n^{-1/2}). \qquad (26)$$

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 (26) 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 D.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 E.3 *Under Assumptions 1, 2 and 3, if $\delta_n \log\log(n) = o(1)$, then*

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

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

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,

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

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 $t_n \in \mathcal{T}_n = \{t \in [-1, 1] : \|t\|_e \leq C n^{-1/2}\}$ 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 E.4 *Under Assumptions 1, 2 and 3, if* $\delta_n \log\log(n) = o(1)$*, then*

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

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 $t_n \le Cn^{-1/2}$ there exists a constant $0 < C_1 < \infty$ such that

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

where the last equality follows from Lemma E.3. ■

Lemma E.5 *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* $t_n \in \mathcal{T}_n$

$$n^{-1}(L_n(\theta(t_n,g)) - L_n(\theta)) =$$

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

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,

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

Next, note that

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

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

$$(3) = \mu_n \{R(\theta(t_n, g), \theta_0) - R(\theta, \theta_0)\} = o_p(n^{-1}) \text{ (see Lemma E.4)}.$$

■

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

Lemma E.6 *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 E.5

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

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

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

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. ■

F. Proof of Theorem 6.1

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 E.2

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

By Lemmas E.5 and E.2

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

Then by Lemma E.3 and properties of the quadratic forms, the maximizer belongs to $\mathcal{T}_n$ and is equal to $\mu_n \{g_n(\theta_{0n})\}$. Hence,

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

As a result,

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

Finally,

$$n^{-1}[L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n)] \geq \sup_{\theta_0 \in \Theta_0 \bigcap \mathrm{NE}} \frac{\|\mu_n \{g_n(\theta_{0n})\}\|_e^2}{2} + o_p(n^{-1}) \tag{27}$$

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 E.6 and Lemma E.3) and by applying Lemma E.5 and

Lemma E.6 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 E.6 to t^* and $g_n(\theta_0)$, I get that

$$n^{-1}[L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n)] \leq \sup_{\theta_0 \in \Theta_0 \bigcap \mathrm{NE}} \frac{\|\mu_n\{g_n(\theta_{0n})\}\|_e^2}{2} + o_p(n^{-1}). \tag{28}$$

It remains to show that such ϵ_n exists. Note that

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

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

Combining (27) and (28) 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 $\alpha_n = o(1)$

$$\int_0^{\alpha_n} \sqrt{\log N_{[]} \left(\epsilon \|g\|_{L^2(P_0)}, \mathcal{G}_n, L^2(P_0) \right)} d\epsilon \leq \int_0^{\alpha_n} K\epsilon^{-\frac{d_x}{\eta + \eta_0}} d\epsilon = o(1).$$

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,

$$\sup_{\|\theta_0-\theta'_0\|_c<\alpha_n} \|g_n(\theta_{0n}) - g_n(\theta'_{0n})\|_{L^2(P_0)} \leq$$
$$\leq \sup_{\|\theta_0-\theta'_0\|_c<\alpha_n} \|g(\theta_0) - g(\theta'_0)\|_{L^2(P_0)} + o(1) \leq O(\alpha_n).$$

G. Proof of Theorem 7.1

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

$$L^w_n(\hat{\theta}^w_n) = \sup_{\theta\in\Theta_{k(n)}} L^w_n(\theta),$$
$$L^w_n(\tilde{\theta}_n) = \sup_{\theta\in\Theta_{k(n)}\bigcap \mathrm{NE}(\hat{\theta}_n)} L^w_n(\theta).$$

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

Lemma G.1 *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* $t_n \in \mathcal{T}_n$

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

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

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

Lemma G.3 *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 6.1.

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 E.2

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

By lemmas G.1 and G.3

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

Then by Lemma E.3 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,

$$n^{-1}[L_n^w(\hat{\theta}_n^w) - L_n(\tilde{\theta}_n^w) \geq \sup_{\theta_0 \in \Theta_0} \frac{\left\|\mu_n \left\{(w-1) g_n(\theta_{0n})\right\}\right\|_e^2}{2} + o_p(n^{-1}) \tag{29}$$

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

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

$$p(\theta^*) = p(\hat{\theta}_n^w) + t_n^* g_n(\theta_{0n}) p_0,$$

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 E.6 and Lemma E.3) and by applying Lemma G.1 and Lemma E.6 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 E.6 to t^* and $g_n(\theta_0)$, I get that

$$n^{-1}[L_n^w(\hat{\theta}_n) - L_n^w(\tilde{\theta}_n)] \leq \sup_{\theta_0 \in \Theta_0} \frac{\|\mu_n\{(w-1)g_n(\theta_{0n})\}\|_e^2}{2} + o_p(n^{-1}). \tag{30}$$

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 (29) and (30) 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 6.1.

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.

H. Pseudo-algorithm for implementation

Computation of the test statistic and the bootstrap critical values involves several constrained nonlinear maximization problems. To compute one of the statistics, say $T_{n,0}$, one needs to maximize the same objective function twice: constrained and unconstrained. These computations can be substantially simplified since for every fixed $\beta \in B$ the objective function is concave and the constrains are affine in the sieve coefficients. Moreover, one does not need to directly compute the finite dimensional-constraints m_j, $j = 0, 1, 2, 3$. Instead, one can evaluate the infinite-dimensional NE constrains ($\tilde{m}_j$, $j = 0, 1, 2, 3$; see equations (19)-(22) in the appendix) at sufficiently large number of points.

Below I present an algorithm for evaluation of the objective function and the constraints at (β, γ), where $\gamma \in \mathbb{R}^{3J}$ is a vector of sieve coefficients and J is the number of known basis functions $\{\psi_j(\cdot,\cdot)\}_{j=1}^J$. Let γ^y be a vector of sieve coefficients that correspond to $h(y,\cdot,\cdot)$. That is, $\gamma = (\gamma^{(0,0)\intercal}, \gamma^{(1,1)\intercal}, \gamma^{(1,0)\intercal})^\intercal$.

Computation of the objective function $\hat{L}_n(\beta, \gamma)$.

1. For every x_i and y_i in the sample
   (i) Compute $A^M(x_i, \beta)$ and $A(y_i, x_i, \beta)$.
   (ii) If $y_i \neq (0, 1)$ compute
   $$P(y_i, x_i, \beta, \gamma) = \int_{A(y_i,x_i,\beta)} f_{e|x}(e|x_i; \beta)de + \sum_{j=1}^{J} \gamma_j^{y_i} \int_{A^M(x_i,\beta)} \psi_j(e, x_i) f_{e|x}(e|x_i; \beta)de.$$
   (iii) If $y_i = (0, 1)$ compute
   $$P(y_i, x_i, \beta, \gamma) = \int_{A(y_i,x_i,\beta)\cup A^M(x_i,\beta)} f_{e|x}(e|x_i; \beta)de - \\ - \sum_{j=1}^{J} (\gamma_j^{(0,0)} + \gamma_j^{(1,1)} + \gamma_j^{(1,0)}) \int_{A^M(x_i,\beta)} \psi_j(e, x_i) f_{e|x}(e|x_i; \beta)de.$$
2. Compute $L_n(\beta, \gamma) = \sum_{i=1}^{n} \log P(y_i, x_i, \beta, \gamma)$.

Computation of the constraints.

1. Create grids $\hat{X}$ and $\hat{U}$ on X and on $[0, 1]^2$, respectively (one can set $\hat{X} = \{x_i\}_{i=1}^n$).
2. For every $x \in \hat{X}$

(i) Compute $A^M(x, \beta)$.

(ii) Transform $\hat{U}$ to a new grid $\hat{E}$ such that $\hat{E}$ is a grid on $A^M(x, \beta)$.

(iii) For every $e \in \hat{E}$

(a) Compute $m_j(x, e, (b, h)),\ j = 0, 1, 2, 3$ using equation (19)-(22) in the appendix. Depending on the statistic we have four equality/inequality constraints evaluated at one point (x, e)

(iv) After repeating step 2.(iii) for every $e \in \hat{E}$ we end up having $4\left\|\hat{E}\right\|$ constraints.

3. After repeating step 2 for every $x \in \hat{X}$ we end up having $4\left\|\hat{E}\right\|\left\|\hat{X}\right\|$ constraints.

For every fixed $\beta \in B$ one can compute a matrix $C(\beta) \in \mathbb{R}^{4\|\hat{E}\|\|\hat{X}\| \times 3J}$ and a vector $c(\beta) \in \mathbb{R}^{4\|\hat{E}\|\|\hat{X}\|}$ such that the constraints are represented as $C(\beta)\gamma \geq (=)c(\beta)$. Thus, computing

$$-\hat{L}_n(\beta) = \min_{C(\beta)\gamma \geq (=)c(\beta)} [-L_n(b, \gamma)]$$

is a convex optimization problem. Finally, one needs to compute

$$L_n(\hat{\theta}) = -\min_{\beta \in B} [-\hat{L}_n(\beta)].$$