A Geometric Approach to Inference in Set-Identified Entry Games

Christian Bontemps[*] Rohit Kumar[†]

December 2019

Abstract

In this paper, we consider inference procedures for entry games with complete information. Due to the presence of multiple equilibria, we know that such a model may be set-identified without imposing further restrictions. We complete the model with the unknown selection mechanism and characterize geometrically the set of predicted choice probabilities, in our case, a convex polytope with many facets. Testing whether a parameter belongs to the identified set is equivalent to testing whether the true choice probability vector belongs to this convex set. Using tools from the convex analysis, we calculate the support function and the extreme points. The calculation yields a finite number of inequalities, when the explanatory variables are discrete, and we characterized them once for all. We also propose a procedure that selects the moment inequalities without having to evaluate all of them. This procedure is computationally feasible for any number of players and is based on the geometry of the set. Furthermore, we exploit the specific structure of the test statistic used to test whether a point belongs to a convex set to propose the calculation of critical values that are computed once and independent of the value of the parameter tested, which drastically improves the calculation time. Simulations in a separate section suggest that our procedure performs well compared with existing methods.

Keywords: set-identification, entry games, convex set, support function

[*]Corresponding author. ENAC & Toulouse School of Economics, Université de Toulouse, 1, Esplanade de l'université, 31080 Toulouse cedex 06, France, email: christian.bontemps@tse-fr.eu.

[†]Indian Statistical Institute (ISI), 7 S.J.S Sansanwal Marg, New Delhi 110016, India, email: sirohi.rohit@gmail.com

1 Introduction

This paper provides an estimation procedure for empirical models of entry and market structure, also called entry games, which may be set-identified. Entry games are very popular in the empirical Industrial Organization literature because they allow researchers to study the nature of firms' profits and the nature of competition between firms from data that are generally easy to collect. They were popularized by the seminal works of Bresnahan and Reiss (1991a), Bresnahan and Reiss (1991b) and Berry (1992).[1] However, the econometric analysis of entry games is complicated by the presence of multiple equilibria, a problem that affects the standard estimation strategy. Without additional assumptions, the model is indeed *incomplete*. Various solutions have been proposed in the literature. First, assumptions can be added regarding the (unknown) selection mechanism in regions of multiple equilibria. Reiss (1996) considers a specific order of entry, and Bjorn and Vuong (1984), randomly draw the equilibrium selection. Bajari et al. (2010) introduce a parametric specification of the selection mechanism, and Grieco (2014) extends it to a non-parametric function of observables and non-observables. Alternatively, it is sometimes possible to estimate an entry game from the observation of some outcome that is independent of the true selection mechanism. In their seminal work with heterogeneous firms, Bresnahan and Reiss (1991b) report that the model uniquely predicts the number of active firms. Berry (1992) generalizes the estimation of the profit function from the observed number of active firms for more than two players. A tremendous number of empirical applications have followed this path (see the survey by de Paula (2012)), notably including Mazzeo (2002), Reiss (1996), Cleeren et al. (2010), and Sampaio (2007), among others. However, adding assumptions could lead to a misspecified model, and working on certain invariant outcomes may lead to a suboptimal procedure (see Tamer (2003)). The recent and blossoming literature on partial/set-identification, following earlier work by Manski (1995), makes it possible to estimate a model that does not uniquely predict actions by using bounds. Following Tamer (2003), the presence of multiple equilibria does not imply partial identification,[2] but the literature provides inference methods that are eligible in both cases. Ciliberto and Tamer (2009)

[1] See Berry and Reiss (2007) for a survey.

[2] However, identification at infinity, as in Tamer (2003), may lead to inference procedures that are non standard; see Khan and Tamer (2010).

were the first to use bounds to estimate an entry game with complete information.

In this paper, we complete the model with the selection mechanism, $\eta(\cdot)$, and characterize the set of predicted choice probabilities generated by the variation of $\eta(\cdot)$ in the space of admissible selection mechanisms. Our first contribution is to characterize more deeply the geometric structure of this set. The set is a convex polytope with many facets (because we focus on pure strategy Nash equilibria), and the number of facets increases exponentially with the number of players. Alternative equilibrium concepts have been proposed in the literature (as in Aradillas-Lopez and Tamer (2008), Beresteanu et al. (2011) or Galichon and Henry (2011)). Changing the equilibrium concept affects some of the calculations provided in this paper and, sometimes, increases the complexity but does not alter the general philosophy. Moreover, Nash equilibrium is the most commonly used solution concept. In this paper, we derive a closed-form expression for the support function of this polytope, the extreme points (or *vertices*) of which can also be calculated as a function of the primitives of the model. These vertices are indeed characterized by an order of outcome selection in the regions of multiple equilibria. Each vertex is also geometrically defined by the intersection of some supporting hyperplanes. We are able to define the cone of outer normal vectors of these hyperplanes and, thereby, the inequalities that are binding in this point.

Testing whether a parameter belongs to the identified set is equivalent to testing whether the true choice probability vector belongs to this convex set. Following Rockafellar (1970), the support function defines a continuum of inequalities that have to be satisfied for any point in the set. This characterization has already been used in the set-identification literature by, in particular, Beresteanu and Molinari (2008), Beresteanu et al. (2011) and Bontemps et al. (2012). Geometrically, the support function in a given direction leads to an inequality that detects whether the point of interest belongs to the same halfspace than the convex set itself. This continuum of inequalities simplifies here toward a finite number because of the specific structure of the entry game considered in this paper.

However, when the number of players increases, the number of facets of the polytope increases exponentially, and, therefore, the smallest number of inequalities necessary to have a sharp characterization - from 16 in a game with three players to more than one million in a game with six players. The standard approach for moment inequality models is to first evaluate all the moments, then eliminate the ones that are far from being binding (see Andrews and Soares, 2010, for the unconditional case). Our second contribution is to provide a geometric selection procedure that does not require us to evaluate all the moments and for which the computational cost is polynomial in the dimension of the outcome space (2^N). The idea is to recursively identify the extreme point of the set that is the most relevant for demonstrating that our point, the true choice probability vector, belongs to the polytope. We then test only the inequalities related to the facets of this extreme point and do not evaluate the other inequalities. The number of relevant moments grows at the rate 2^N as in Ciliberto and Tamer (2009), but with a sharper characterization of the set. This is a considerable improvement in terms of computational burden with respect to alternative sharp characterization methods (such as the two step approaches of Andrews and Soares (2010) and Romano et al. (2014)).

Furthermore, and more importantly, we develop a test procedure for the hypothesis of whether the true choice probability vector belongs to this polytope that exploits the specific structure of the test statistic. When we test whether a parameter θ belongs to the identified set, the feasible test statistic depends on the support function of the polytope, which is fixed, and the choice probability vector, which is estimated. The asymptotic distribution depends on θ in a specific manner. It allows us to derive critical values for our test procedures that are calculated once and valid for any θ. This is a tremendous simplification with respect to a general moment inequality model. Some of the recommended test procedures can be conservative, but, one, which exploits the maximum number of facets at any extreme point, is not. This maximum number can be determined by brute force using the geometric structure of the set; we also provide an upper bound. Simulations highlight that our method works well with the sample sizes that are usually considered in the empirical literature (from 500 to 2000 observations) and outperforms existing inference procedures.

This paper belongs to the growing literature on set-identification and lies at the intersection between the moment inequality literature and the literature on set-identification that exploits the geometric structure. A model is generally set-identified when the data are incomplete because of some missing information (a censorship mechanism, two-sample combination, or multiple equilibria). In entry games, the unknown selection mechanism in regions of multiple equilibria is the missing one, but it is naturally bounded between 0 and 1. These bounds lead to moment inequalities (see, for example, Ciliberto and Tamer (2009)). Moment inequality procedures are now well developed in the partial identification literature, which includes contributions by, among others, Chernozhukov et al. (2007), Rosen (2008), Andrews and Soares (2010), Bugni (2010), Canay (2010), Romano and Shaikh (2008), Romano and Shaikh (2010), Chernozhukov et al. (2015), Andrews and Shi (2013) or Aradillas-Lopez and Rosen (2017). However, in most cases, the structure of the missing information is not exploited. Other contributions exploit this missing information by characterizing the identified set directly or indirectly through a convex set, and, then, through the support function. The support function provides a collection of moment inequalities, but these moment inequalities have a particular structure and are exploited differently than in the general moment inequality literature (see, in particular, the two surveys by Molchanov and Molinari (2015) and Bontemps and Magnac (2017)).

Papers that have exploited the structure of the missing information either use random set theory (introduced in econometrics by Beresteanu and Molinari, 2008), optimal transport (Galichon and Henry (2011)) or complete the model, as in Bontemps et al. (2012). Fundamentally, all methods are intended to provide a sharp characterization of the identified set and lead to a collection of moment inequalities that are specifically determined using the tools of the approach considered. They all lead to the same set of inequalities. In this spirit, they can be seen as equivalent. Redundant ones are eliminated through the characterization of what Galichon and Henry (2011) call the core determining class. In our case, we also characterize the core determining class, i.e. the collection of facets of the convex set, by a necessary and sufficient condition. However, in addition to these equivalences, our geometric approach is able to select the subsets of moments that are close to binding without having to evaluate all of them and to determine the maximum number of moments that are binding.

Section 2 considers the general entry game with N players and no explanatory variable. Explanatory variables do not play a specific role in the procedure. We characterize the regions with multiple equilibria and compute the support function of the polytope, i.e., the set of predicted choice probability vectors for a given parameter value. The sequence of inequalities that need to be verified is characterized. The aim of Section 3 is to more deeply exploit the geometry of the polytope to propose a more efficient estimation procedure. We first provide a necessary and sufficient condition for eliminating any redundant inequalities. We then propose our geometric selection procedure which consists of determining the local extreme point of the polytope and only testing the inequalities relative to the facets at this extreme point. Section 4 calculates the asymptotic distribution of the test statistic used and proposes different calculations of the critical value that are computed once for all. Section 5 proposes a few Monte Carlo simulations to assess the performance of our inference procedure and compare it with existing procedures. Section 6 considers the case with discrete explanatory variables, and Section 7 concludes the paper. The Supplemental Material available online[3] contains appendices that provide the proofs and the algorithms (Appendix A), the specific details for the three player game (Appendix B) and an additional Monte Carlo analysis (Appendix C).

2 Entry game with N players

We formalize the entry game with N players/firms. For exposition purposes, we first consider a model without explanatory variables and then extend it in section 6. We first define some notations before characterizing the identified set.

2.1 Setup and notations

2.1.1 The model

Let N denote the total number of firms that can enter any market. Following Berry (1992), we introduce a model of market structure where the profit function π_{im} of firm i in a market m is assumed to be independent of the identity of the firm's competitors. All firms decide simultaneously whether to enter the market (the action is $a_{im} = 1$) if, in a pure strategy Nash equilibrium, their profit is positive ($\pi_{im} > 0$). Otherwise, $\pi_{im} <= 0$, and the action is $a_{im} = 0$. The profit function is assumed, without loss of generality, to be linear in the explanatory variables[4]

$$\begin{aligned} \pi_{im} &= \beta_i + \alpha_i \left(\sum_{j \neq i} a_{jm} \right) + \varepsilon_{im}, \\ a_{im} &= \mathbf{1}\{\pi_{im} > 0\}. \end{aligned} \tag{1}$$

[3]It can be downloaded, jointly with the full text, by following the link: https://www.tse-fr.eu/sites/default/files/TSE/documents/doc/wp/2018/wp_tse_943.pdf

[4]Any (separable) parametric form $\pi_{im} = f_i\left(\sum_{j\neq i} a_{jm}; \alpha\right) + \varepsilon_{im}$ can be considered as long as the function $f_i(\cdot;\theta)$ is strictly decreasing in its first argument.

Following the literature on entry games, we assume that $\alpha_i < 0$, i.e., the presence of more competitors decreases a firm's profit.[5] The unobserved components ε_{im}, $i = 1, \ldots, N$ are drawn from a known distribution (up to some parameter vector γ). The econometrician does not observe their values, whereas all firms within a market observe them before deciding whether to enter (complete information case).

For identification, we first need a scale normalization, and thus, we assume that the variance of each shock ε_{im} is equal to unity. We denote by $F(\cdot;\gamma)$ the cumulative distribution function of $\varepsilon_m = (\varepsilon_{1m}, \ldots, \varepsilon_{Nm})^\top$, and we assume that the distribution is continuous with full support. Henceforth, we use the notation θ for all the parameters in the model ($\theta \in \Theta \subseteq \mathbb{R}^l$), [6] and we omit the subscript m for notational convenience.

2.1.2 The multiplicity of pure strategy Nash equilibria

For a given market, an outcome y is the vector of actions (in $\{0,1\}^N$) taken by the firms. There are obviously 2^N possible outcomes from $(0,\ldots,0)^\top$ to $(1,\ldots,1)^\top$. We denote by $\mathcal{Y}$ this set of possible outcomes. $\mathcal{Y}_K$ denotes the subset of outcomes with K active firms in equilibrium, i.e. any K firms playing action 1. There is 1 outcome with 0 active firms, N outcomes with 1 active firm and $d_K = \binom{N}{K}$ outcomes with K active firms for $K \leq N$. For each K, we label the outcomes as $y_j^{(K)}$, $j = 1, \ldots, d_K$ according to a predefined order.[7] Globally, we order the outcomes in $\mathcal{Y}$ first by their number of active firms, then according to the predefined order within each $\mathcal{Y}_K$:

$$\mathcal{Y} = \left\{ \underbrace{y_1^{(0)}}_{\mathcal{Y}_0}, \underbrace{y_1^{(1)}, \ldots, y_{d_1}^{(1)}}_{\mathcal{Y}_1}, \ldots, \underbrace{y_1^{(K)}, \ldots, y_{d_K}^{(K)}}_{\mathcal{Y}_K}, \ldots, \underbrace{y_1^{(N)}}_{\mathcal{Y}_N} \right\}.$$

It is well known that the model has multiple equilibria, i.e., there are regions of realizations of ε in which we cannot uniquely predict each firm's action. Consequently, there is no one-to-one mapping between the collection of possible outcomes and the regions of ε given any parameter value θ.

What is missing from the model is the selection of a given equilibrium in the regions of multiple equilibria. We define this selection mechanism $\eta(\cdot)$ as in Definition 2 of Galichon and Henry (2011).

[5] The case in which $\alpha_i > 0$ could be handled equivalently. Gualdini (2018) considers, in a network formation game, these two cases.

[6] $\theta = (\beta_1, \ldots, \beta_N, \alpha_1, \ldots, \alpha_N, \gamma)^\top$

[7] The order for $K = 1$ is $((1,0,\ldots,0)^\top$, $(0,1,0\ldots,0)^\top$ ... $(0,\ldots,0,1)^\top)$, and so forth.

Definition 1 (Equilibrium selection mechanism) *An equilibrium selection mechanism is a conditional probability* $\eta(\cdot|\varepsilon;\theta)$ *of* y *given* ε *such that the selected value of the outcome variable is actually an equilibrium predicted by the game.*

We denote by $\mathcal{E}$ the set of selection mechanisms and by $P(\theta,\eta)$ the predicted choice probability vector when the parameter of the model is θ and the selection mechanism is $\eta(\cdot)$. We partition this vector according to the partition of $\mathcal{Y}$ as[8]

$$P(\theta,\eta) = \Bigg(\underbrace{P_1^{(0)}(\theta,\eta)}_{P^{(0)}(\theta,\eta)},\ldots,\underbrace{P_1^{(K)}(\theta,\eta),\ldots,P_{d_K}^{(K)}(\theta,\eta)}_{P^{(K)}(\theta,\eta)},\ldots,\underbrace{P_1^{(N)}(\theta,\eta)}_{P^{(N)}(\theta,\eta)}\Bigg)^{\top}. \tag{2}$$

One solution to the multiple equilibria problem consists of making assumption on this selection mechanism like in Reiss (1996) or Cleeren et al. (2010), for example. The vector of predicted choice probabilities is a point in $[0,1]^{2^N}$ and standard inference techniques can be used. This is, of course, ad hoc and may lead to misspecification.

Another solution, ours, following the literature on set-identification, consists of characterizing all the possible choice probabilities predicted by the model. The vector of predicted choice probabilities, instead of being a point, belongs to a (convex) set that we characterize. Different sets of values (θ,η) may generate the same point $P(\theta,\eta)$.[9] Our goal is to characterize the ones which generate the true choice probability vector. In the next subsection, we first characterize the set of choice probabilities predicted by the model.

2.2 From the set of choice probabilities to the identified set

In this section, we want to characterize the set of predicted choice probabilities. To do so, we need to understand the multiplicity structure and characterize it. Then, we derive a parametrization of the set.

[8] In particular, $P_i^{(K)}(\theta,\eta)$ denotes $Prob(y = y_i^{(K)}|\theta,\eta)$.

[9] Observe that having multiple equilibria does not automatically guarantee to have set/partial identification. In the following, our statistical procedure, which consists of inverting a test, is valid for point or set-identified models.

2.2.1 The regions of multiple equilibria

Our specification ensures that multiple equilibria only involve outcomes with the same number of active firms, i.e., within $\mathcal{Y}_K$.[10] Therefore, we focus on subsets of outcomes $S \subseteq \mathcal{Y}_K$ to characterize the multiple equilibria regions. We say that a subset $S \subseteq \mathcal{Y}_K$ **is in multiplicity** if the prediction of the game is all outcomes in S and no outcome outside S for ε in a non empty set, $\mathcal{R}_S^{(K)}(\theta)$. $\mathcal{R}_S^{(K)}(\theta)$ is called a multiple equilibria region. We denote by $S^{(K)}$ the collection of subsets S of $\mathcal{Y}_K$ in multiplicity.[11]

$$S^{(K)} = \Big\{ S \subseteq \mathcal{Y}_K : |S| \geq 2 \text{ and } S \text{ is in multiplicity} \Big\}.$$

Note that not all subsets of cardinality greater than two are elements of $S^{(K)}$. For example, when $N = 4$ and $K = 2$, $S_1 = \big\{(1,1,0,0)^\top, (0,0,1,1)^\top\big\}$ is not in multiplicity whereas the subset $S_2 = \big\{(1,1,0,0)^\top, (1,0,1,0)^\top\big\}$ is.

We now present a necessary and sufficient condition for S to be in multiplicity. Define N_0 (resp. N_1) the set of indices of firms that always play action 0 (resp. 1) across S. n_0 and n_1 are their cardinalities. N_0 and N_1 being fixed, there are $\binom{N-n_0-n_1}{K-n_1}$ possible outcomes in $\mathcal{Y}_K$ corresponding to the remaining choice of the $K - n_1$ firms which play action 1 among the $N - n_0 - n_1$ remaining ones. S should contain all these possibilities to be in multiple equilibria and it is formalized in the next proposition.

Proposition 1 *A set $S \subseteq \mathcal{Y}_K$ is in multiplicity if and only if $|S| = \binom{N-n_0-n_1}{K-n_1}$.*

Observe that, for our particular examples above, S_1 is not in multiplicity because $n_0 = n_1 = 0$ and, consequently, the subset should contain $\binom{4}{2} = 6$ outcomes with two active firms to be in multiplicity. S_2 is in multiplicity because $n_0 = n_1 = 1$ and it collects all possible outcomes $(\binom{4-1-1}{2-1} = 2)$. The proof of Proposition 1 characterizes also the region $\mathcal{R}_S^{(K)}(\theta)$ of ε.

Following Proposition 1, we count the number of multiple equilibria regions.

[10] The proof of Berry (1992) can be easily transposed. See the Supplemental Material.

[11] Note that the maximum number of such subsets is equal to $2^{d_K} - d_K - 1$.

Proposition 2 *The cardinality of* $S^{(K)}$*, i.e., the number of multiple equilibria regions predicting* K *active firms, for* $1 \leq K \leq N-1$ *is equal to*

$$\left|S^{(K)}\right| = \sum_{n_1=0}^{K-1} \sum_{n_0=0}^{N-K-1} \binom{N}{n_1}\binom{N-n_1}{n_0}.$$

When $K=1$, the number of regions of multiple equilibria is $\sum_{n=2}^{N}\binom{N}{n}$, i.e., all possible combinations of more than two outcomes. However, as illustrated in Table I, the number of regions for various values of N and K is generally far less from all the possible combinations. It means that the parametrization of the set of predicted choice probabilities is of a much lower dimension than one would have expected.

[**Include Table I**]

2.2.2 The set of predicted choice probabilities

We also define the subset of $S^{(K)}$ that contains one specific outcome $y_j^{(K)}$ as

$$S_j^{(K)} = \left\{ S \in S^{(K)} : \, y_j^{(K)} \in S \right\}.$$

Following Berry and Tamer (2007) and Galichon and Henry (2011), we can calculate the probability of observing outcome $y_j^{(K)}$. This probability depends on the parameter vector θ and on the unknown selection mechanism η that selects equilibrium $y_j^{(K)}$ in the regions of multiple equilibria that predicts this outcome. More specifically,

$$P_j^{(K)}(\theta,\eta) = \int_{U_j^{(K)}(\theta)} dF(\varepsilon;\gamma) + \sum_{S \in S_j^{(K)}} \int_{\mathcal{R}_S^{(K)}(\theta)} \eta\big(y_j^{(K)}|\varepsilon;\theta\big) dF(\varepsilon;\gamma), \tag{3}$$

where $U_j^{(K)}(\theta)$ is the region of $\varepsilon \in \mathbb{R}^N$ which uniquely predicts the outcome $y_j^{(K)}$. Let us denote by

$$\Delta_j^{(K)}(\theta) = \int_{U_j^{(K)}(\theta)} dF(\varepsilon;\gamma) \;\text{ and }\; \Delta_S^{(K)}(\theta) = \int_{\mathcal{R}_S^{(K)}(\theta)} dF(\varepsilon;\gamma) \;\text{ for } S \in S^{(K)}.$$

Let $A(\theta)$ (resp. $B_K(\theta)$, for any $K=0,\ldots,N$) be the set of $P(\theta,\eta)$ (resp. $P^{(K)}(\theta,\eta)$) generated by the variation of η in $\mathcal{E}$

$$A(\theta) = \left\{ P \in \mathbb{R}^{2^N} : \, \exists \eta \in \mathcal{E}, \; P = P(\theta,\eta) \right\}, \; B_K(\theta) = \left\{ P_K \in \mathbb{R}^{d_K} : \, \exists \eta \in \mathcal{E}, \; P_K = P_K(\theta,\eta) \right\}.$$

Equation (3) is a parametrization of the sets $A(\theta)$ and $B_K(\theta)$, $K = 0, \ldots, N$. This parametrization is indexed by the regions $\mathcal{R}_S^{(K)}(\theta)$, counted in Table I.

2.2.3 A characterization of the identified set

Let $P_0 = P(\theta_0, \eta_0)$ be the true choice probabilities generated by the true (unknown) parameter θ_0 and the true (unknown) selection mechanism η_0. The identified set Θ_I is defined as the collection of points θ such that P_0 can be rationalized with a selection mechanism

$$\Theta_I = \Big\{\theta \in \Theta : \text{ such that } \exists \eta \in \mathcal{E},\ P_0 = P(\theta, \eta)\Big\}. \tag{4}$$

The following is easily verified:

$$\theta \in \Theta_I \text{ if and only if } P_0 \in A(\theta). \tag{5}$$

We therefore need to be more precise about the structure of $A(\theta)$ to be able to verify the second part. The following result holds:

Proposition 3 *$A(\theta)$ is a convex set of $\mathbb{R}^{2^N}$, and*

$$A(\theta) = B_0(\theta) \times B_1(\theta) \times B_2(\theta) \times \ldots \times B_N(\theta),$$

where $B_K(\theta)$ is a convex set in $\mathbb{R}^{d_K}$.

The convexity of $A(\theta)$ is a general feature of an entry game and does not depend on our specification (see Beresteanu et al. (2011)). Its specific structure, i.e., the direct product of several components, comes from our specification in Equation (1) which ensures the unicity of the number of active firms in the regions of multiple equilibria. This structure simplifies some of the following results of this section.

Also, $B_K(\theta)$ is a point only when the number of active firms in equilibrium is 0 or N, because there is no region of multiple equilibria involving these specific outcomes. Note that each $B_K(\theta)$ is strictly included[12] in the cube, Cub_K, defined by

$$\Delta_j^{(K)}(\theta) \leq P_j^{(K)} \leq \Delta_j^{(K)}(\theta) + \sum_{S \in S_j^{(K)}} \Delta_S^{(K)}(\theta),\ \forall j = 1, \ldots, d_K \tag{6}$$

[12]See Appendix B, Figure 7, for a visual illustration of the case with three players and $K = 1$.

It comes from the fact that $\eta(\cdot)$ in Equation (3) lies between 0 and 1.

Θ_I, the identified set, is not convex, but it can be characterized by verifying that a point, P_0, belongs to a convex set, $A(\theta)$. Using Proposition 3, we can decompose this condition into the following sub-conditions:

$$P_0 \in A(\theta) \text{ iff } \forall K \in \{0, 1, \ldots, N\},\ P_0^{(K)} \in B_K(\theta).$$

2.3 The support function and a first selection of moment inequalities

Following the convex literature, we introduce the support function of each convex set $B_K(\theta)$. This tool has, in particular, been used, in the set-identified literature, by Beresteanu and Molinari (2008) and Bontemps et al. (2012). It helps in generating the set of inequalities satisfied by P_0. We first recall what the support function of a convex set is and how it generates the inequalities that are the basis of our inference procedure. The *support function* of a convex set $A \subset \mathbb{R}^d$ is defined as:

$$\delta^*(q; A) = \sup_{x \in A} \ q^\top x.$$

for all directions, $q \in \mathbb{R}^d$. Its geometrical interpretation is illustrated in Figure 1. The support function of a convex set in a given direction locates the supporting hyperplane in this direction. For each direction q, it defines an inequality that is satisfied by any point of the convex set. The support function implicitly gathers all the inequalities that define the convex set into a single function. If the set is smooth, there is a continuum of such inequalities; if it is a polytope, only a finite number of inequalities is necessary to characterize the set. Kaido and Santos (2014) show that, when the set is convex, using the support function leads to an efficient estimator of the convex identified set.

Following Rockafellar (1970) and Proposition 3, the identified set is characterized by the following inequalities

$$\begin{aligned} \theta \in \Theta_I &\Longleftrightarrow P_0 \in A(\theta) \\ &\Longleftrightarrow \forall q \in \mathbb{R}^{2^N},\ q^\top P_0 \le \delta^*(q; A(\theta)), \\ &\Longleftrightarrow \forall K,\ P_0^{(K)} \in B_K(\theta), \\ &\Longleftrightarrow \forall K,\ \forall q_K \in \mathbb{R}^{d_K},\ q_K^\top P_0^{(K)} \le \delta^*(q_K; B_K(\theta)). \end{aligned} \tag{7}$$

We now turn to the calculation of the support function of $B_K(\theta)$ for any K. Let $q_K \in \mathbb{R}^{d_K}$ be a given direction. We assume the following order among the coordinates of q_K: $q_{i_1,K} \ge q_{i_2,K} \ge$

$\ldots \geq q_{i_{d_K},K}$. We also partition $S^{(K)}$, the collection of subsets of outcomes with K active firms in multiplicity, as follows: we denote $\mathcal{O}_{i_1}^{(K)} = S_{i_1}^{(K)}$, the elements of $S^{(K)}$ which contain the outcome $y_{i_1}^{(K)}$ and by $\mathcal{O}_{i_2}^{(K)}$ the subset of elements of $S_{i_2}^{(K)}$ that are not in $\mathcal{O}_{i_1}^{(K)}$, i.e., $S_{i_2}^{(K)} \backslash S_{i_1}^{(K)}$ and more generally $\mathcal{O}_{i_j}^{(K)} = S_{i_j}^{(K)} \backslash \cup_{k<j} S_{i_k}^{(K)}$, for any $j \leq d_K$. Note that the construction of the outcomes $\mathcal{O}_j^{(K)}$'s is linked to the order of the components of q_K. We now provide a closed-form expression for the support function in this direction.

Proposition 4 *Let $q_K \in \mathbb{R}^{d_K}$, and assume $q_{i_1,K} \geq q_{i_2,K} \geq \ldots \geq q_{i_{d_K},K}$. The support function in the direction q_K, $\delta^*(q_K; B_K(\theta))$, is equal to:*

$$\delta^*(q_K; B_K(\theta)) = \sum_{j=1}^{d_K} q_{j,K} \Delta_j^{(K)}(\theta) + \sum_{j=1}^{d_K} q_{i_j,K} \left(\sum_{S \in \mathcal{O}_{i_j}^{(K)}} \Delta_S^{(K)}(\theta) \right). \tag{8}$$

It is reached at the extreme point

$$E_{i_1,i_2,\ldots,i_{d_K}}^{(K)} = \operatorname{vec}\Big(\Delta_1^{(K)}(\theta) + \sum_{S \in \mathcal{O}_1^{(K)}} \Delta_S^{(K)}(\theta), \ldots, \Delta_{d_K}^{(K)}(\theta) + \sum_{S \in \mathcal{O}_{d_K}^{(K)}} \Delta_S^{(K)}(\theta) \Big).$$

Consequently, $B_K(\theta)$ is a polytope, and its vertices are included in the set of points $E_{i_1,i_2,\ldots,i_{d_K}}^{(K)}$ where the vector of indices $(i_1, \ldots, i_{d_K})$ is any permutation of the vector of indices $(1, 2, \ldots, d_K)$. $B_K(\theta)$ has at most $d_K!$ vertices.

Each extreme point of $B_K(\theta)$, and therefore its support function, can be calculated from the knowledge of the non-zero values of $\Delta_S^{(K)}(\theta)$, $S \in S^{(K)}$. This number of non-zero values is the number of multiple equilibria regions and we saw in Proposition 2 that this number is much smaller than $2^{d_K} - d_K - 1$ (see table I). Consequently, the parametrization of $B_K(\theta)$ is numerically tractable for moderate values of N. Furthermore, each non-zero value $\Delta_S^{(K)}(\theta)$ can easily be calculated or simulated from the knowledge of the distribution of ε.

We can now extend this result to the calculation of the support function of the full set $A(\theta)$ for any direction $q \in \mathbb{R}^{2^N}$. We adopt the standard notation: $q = \operatorname{vec}(q_0, q_1, \ldots, q_N)$, where q_K is the direction related to the set $B_K(\theta)$ (i.e., $q_K \in \mathbb{R}^{d_K}$) and $\operatorname{vec}(\cdot)$ denotes the vertical concatenation.

Proposition 5 *The support function of $A(\theta)$ in direction q is equal to*

$$\delta^* (q; A(\theta)) = \sum_{K=0}^{N} \delta^* (q_K; B_K(\theta)) . \tag{9}$$

This results come from the specific characterization of $A(\theta)$ in Proposition 3. The last proposition, combined with Equation (7), is the basis of our inference procedure. It generates a continuum of inequalities that have to be satisfied for any parameter of the identified set. However, since all the $B_K(\theta)$'s and, therefore, $A(\theta)$, are polytopes, it is necessary and sufficient to test the inequalities in a finite set of directions. We now explicit this set of directions, first for the $B_K(\theta)$'s than for $A(\theta)$.

Let $\mathcal{Q}_K$ be the set of non-null directions of $\mathbb{R}^{d_K}$ with coordinates that are either one or zero. There are $2^{d_K} - 1$ directions in $\mathcal{Q}_K$. The next proposition shows that it is sufficient to check the inequalities in $\mathcal{Q}_K$, for all K, to characterize the identified set.

Proposition 6

$$\theta \in \Theta_I \iff \forall K \in \{0, 1, 2, \ldots, N\}, \quad \forall q_K \in \mathcal{Q}_K,\, q_K^\top P_0^{(K)} \leq \delta^*(q_K; B_K(\theta)).$$

Remark We already mentioned that our specification ensures that the number of firms entering the market is constant among outcomes in multiplicity. As a result, the sets $B_K(\theta)$ belong to an hyperplane because the sum of the components of $P^{(K)}(\theta, \eta)$ is a constant which depends on θ only. If we wanted to characterize one $B_K(\theta)$ only, for one specific choice of K, we would need to consider all the directions of $\mathcal{Q}_K$ combined with the direction $(-1, -1, \ldots, -1)$ to ensure the equality of the sum of all components. Here, due to the fact that we are considering $A(\theta)$, which is included in the simplex (the sum of all the probabilities is equal to 1),[13] we don't need to consider this direction. As a matter of fact, if all the inequalities are satisfied with $(1, \ldots, 1)$ for all K, they are equalities and therefore are automatically satisfied for $(-1, -1, \ldots, -1)$.

Optimal transport, random sets or completion of the model Our approach consists in characterizing the set $A(\theta)$ (or, equivalently, the sets $B_K(\theta)$) through its support function and

[13]We thank one of the referees for highlighting this issue.

extreme points. This is done after having completed the model with the unknown selection mechanism, $\eta(\cdot)$, and finding which selection mechanisms generate the extreme points. The geometric structure induced by the multiplicities allows us to exhibit the inequalities that are satisfied by any parameter of the identified set.

Galichon and Henry (2011) use the optimal transport theory and the notion of core determining class to generate the relevant inequalities that characterize sharply the identified set. Beresteanu et al. (2011) emphasize that an entry game is a model with convex predictions. They use random set theory and, in particular, the Aumann expectation considered in their paper is our set $A(\theta)$. Both methods are numerically challenging for a game with 6 players even when considering only pure strategy equilibria. Following Proposition 6, there are, at maximum, $\sum_{K=0}^{N}(2^{d_K}-1)$ inequalities. However, this number is very large when $N \geq 6$; we have more than 1 million of inequalities to check for 6 players.

Ciliberto and Tamer (2009) bound the sets $B_K(\theta)$ by the cubes Cub_K introduced above, which are easier to characterize. Their approach can handle games with a moderate number of players above 6, like our method, but bounding component by component makes the estimated set larger, and sharpness is not attained.

Fundamentally, whether one uses random set theory and the capacity functional, the optimal transport approach of Galichon and Henry (2011) or the approach presented in this paper, all these methods are intended to derive a sufficient set of inequalities satisfied by the parameters in a specific manner. Each method has its specificities. However, our approach allows us to go deeper into the geometric analysis of the set $A(\theta)$ and this is the objective of the next section.

3 Using the geometry of $A(\theta)$ to select inequalities

The convex set $B_K(\theta)$ can be characterized by at most $2^{d_K}-1$ inequalities. Due to its particular geometry, it may be the case that some of these inequalities are redundant. In this section, we present two strategies to reduce the number of inequalities. The first consists of calculating a core determining class introduced by Galichon and Henry (2011) and later used in Chesher and Rosen (2017). The second consists of exploiting further the geometry to propose a geometric selection procedure of the inequalities without having to evaluate all of them.

3.1 Deriving a core determining class of an entry game

The core determining class introduced by Galichon and Henry (2011) yields to a collection of irredundant moment inequalities that are sufficient to sharply characterize the identified set Θ_I. We provide a characterization of the core determining class in an entry game from the geometric study of the multiplicity structure of the model. For the text to be self-contained, we borrow some definitions and concepts from Galichon and Henry (2011).

Definition 2 (Choquet capacity) *A Choquet capacity $\mathcal{L}$ on the set $\mathcal{Y}$ is a set function $\mathcal{L} : C \subseteq \mathcal{Y} \to [0,1]$ that is*

- *normalized, i.e., $\mathcal{L}(\emptyset) = 0$ and $\mathcal{L}(\mathcal{Y}) = 1$, and*
- *monotone, i.e., $\mathcal{L}(C) \leq \mathcal{L}(B)$, for any $C \subseteq B \subseteq \mathcal{Y}$.*

Definition 3 *The smallest class Ω of subsets of $\mathcal{Y}$ is called core determining for the Choquet capacity $\mathcal{L}$ on $\mathcal{Y}$ if $\mathbb{P}(C) \leq \mathcal{L}(C)$ holds for all $C \in \Omega$; then, $\mathbb{P}(C) \leq \mathcal{L}(C)$ holds for all $C \subseteq \mathcal{Y}$.*

The set $A(\theta)$ is characterized by its support function. Thus, we define the Choquet capacity for a subset $C_K \subseteq \mathcal{Y}_K$ as

$$\mathcal{L}(C_K) = \delta^*\big(e_{C_K}; B_K(\theta)\big) = \max_{\eta \in \mathcal{E}} \left(\sum_{j | y_j^{(K)} \in C_K} P_j(\theta, \eta) \right), \tag{10}$$

where $e_{C_K} \in \{0,1\}^{d_K}$ with $\big(e_{C_K}\big)_j = 1$ if $y_j^{(K)} \in C_K$ and 0 otherwise. For a collection of subsets $C = \big\{C_K \subseteq \mathcal{Y}_K : K \leq N\big\}$, the Choquet capacity is defined as $\mathcal{L}(C) = \sum_{k=0}^{N} \mathcal{L}\big(C_K\big)$. $\mathcal{L}$ is monotone, as it is the sum of quantities that are positive and $\mathcal{L}(\mathcal{Y}) = 1$.

We define the concept of connectedness, which is useful for the exposition, introduced by Galichon and Henry (2011). For a subset $C_K \subseteq \mathcal{Y}_K$, we define the (undirected) graph generated by C_K as $\Gamma_{C_K} = (C_K, E)$,[14] where there is an edge between two vertices if they are in multiplicity with eventually some additional outcomes that are only in C_K (no outcome from $\mathcal{Y}_K \setminus C_K$). For any graph $\Gamma = (V, E)$, we say that $C \subseteq V$ is **connected in the graph** Γ if there is a path of elements of E connecting any pair of nodes of C.

[14]Recall that an undirected graph $\Gamma = (V, E)$ is a collection of vertices/nodes V and edges/links E that link these vertices. A graph Γ is connected if any pair of vertices are connected in Γ.

Definition 4 (Well connectedness) *A subset $C_K \subseteq \mathcal{Y}_K$ is called well connected in $\mathcal{Y}_K$ if $\mathcal{Y}_K \backslash C_K$ is connected in the graph $\Gamma_{\mathcal{Y}_K \backslash C_K}$.*

Note that $\mathcal{Y}_K$ is in multiplicity. Therefore, the graph $\Gamma_{\mathcal{Y}_K}$ is connected, and every $C_K \subseteq \mathcal{Y}_K$ is connected in the graph $\Gamma_{\mathcal{Y}_K}$. The notion of well connectedness extends the notion of connectedness by imposing restrictions on the complementary of C_K.

Note that the graph $\Gamma_{\mathcal{Y}}$ is not connected, as there is no multiplicity between $\mathcal{Y}_K$ and $\mathcal{Y}_{K'}$, for $K \neq K'$. Thus, $\Gamma_{\mathcal{Y}_K}$ is a component of $\Gamma_{\mathcal{Y}}$.[15] We collect all well-connected subsets of $\mathcal{Y}_K$ as

$$\Omega_K = \Big\{ C_K \subseteq \mathcal{Y}_K : C_K \text{ well connected in } \mathcal{Y}_K \Big\}$$

Galichon and Henry (2011) present some models in which the core determining class can be of much lower cardinality than $2^{|\mathcal{Y}|}$ by exploiting the monotonicity property in certain supermodular games. However, their approach does not provide a way to find a core determining class for a general entry game. Chesher and Rosen (2017) provide a sufficient condition to characterize a core determining class of set-identified models that can be written into what they call a generalized instrument variable model. Our next proposition provides a complete characterization of a core determining class for our entry model through a necessary and sufficient condition.

Proposition 7 *A collection Ω of subsets of $\mathcal{Y}$ is core determining for $\mathcal{L}$ in (10) if and only if $\Omega = \big\{ \Omega_K : K = 0, 1, \ldots, N \big\}$.*

Subsection A.7 in the Supplemental Appendix provides an algorithm to construct the core determining class from Proposition 7. It is applied for the particular examples of $N = 4$ and $K = 2$ in Subsection A.7.2. However, it does not significantly reduce the number of irredundant moment inequalities in our entry game. For example, when $N = 6$ and $K = 3$, it eliminates fewer than $30,000$ inequalities from a total of $2^{20} - 1 = 1,048,575$.

3.2 A geometric selection procedure

The core determining class is a useful concept because we eliminate redundant inequalities. However, it does not significantly reduce the number of inequalities in our entry game. We now present a

[15]Recall that for an undirected graph $\Gamma = (V, E)$, components of Γ are subgraphs $\{H_i\}_{i=1}^{k}$ such that each H_i is connected and H_i is not connected to H_j for $i \neq j$.

geometric selection procedure that fully exploits the geometry of the sets $B_K(\theta)$. The procedure first selects the extreme point of the set that seems the closest to the vector $P_0^{(K)}$ and then evaluates only the inequalities associated with this extreme point, i.e., tests the directions that are the outer normal vector of the supporting hyperplanes of $B_K(\theta)$ at this point, and this, for each $K = 0, \ldots, N$. Following Proposition 4, an extreme point is determined by an order in the coordinates (note that, a priori, two different orders could lead to the same physical point). The first part of the algorithm is intended to determine this order in a recursive manner by exploiting the position of $P_0^{(K)}$ with respect to the cube Cub_K which contains $B_K(\theta)$. We explain the steps in non-technical detail below and formalize the algorithm in the Supplemental Appendix (Appendix A.8).

Local moment selection Our local moment selection procedure can be summarized as follows:

1. Determine the cube Cub_K that contains $B_K(\theta)$ by calculating the minimum and maximum of each coordinate. Then, determine which coordinate of $P_0^{(K)}$ is the furthest from the center of the cube.

2. Assume this is the j^{th} coordinate.

   (a) If it is on the maximum side, the extreme point is of type $E_{j,?,\ldots,?}^{(K)}(\theta)$, and we now have to determine the next component. To do so, we project $P_0^{(K)}$ on the face, and we repeat the previous calculation by taking into account that we are on the face that maximizes the j^{th} coordinate.

   (b) If it is on the minimum side, we know that our extreme point will be of the form $E_{?,\ldots,?,j}^{(K)}(\theta)$, and we now have to determine the next component. To do so, we project $P_0^{(K)}$ on this face, and we repeat the previous calculation by taking into account that we are on the face that minimizes the j^{th} coordinate.

3. Repeat the following steps until having found one order of coordinates.

4. Once the local extreme point $E_{i_1,i_2,\ldots,i_{d_K}}^{(K)}$ is determined, we can focus on the directions of the local supporting hyperplanes. Let the d_K directions, e_{i_1}, e_{i_1,i_2},..., $e_{i_1,i_2,\ldots,i_{d_1-1}}$, $e_{i_1,i_2,\ldots,i_{d_1}}$, where the components are equal to 1 when the indices are subscripts of e and 0 otherwise. This

set of directions is included in the set of directions of the local supporting hyperplanes. Only checking these directions doesn't provide a sharp characterization of $B_K(\theta)$ unless $K=1$ or $N-1$ but, however, provides an important refinement with respect to the existing method of Ciliberto and Tamer (2009). An algorithm is provided in the Supplemental Appendix to derive the other directions to consider.[16]

Our procedure selects which moments among the $2^{d_K}-1$ are potentially binding without having to evaluate all of them. The selection is based on the spatial location of the point $P^{(K)}$ and exploits the geometry of the set $B_K(\theta)$. Proposition 8 shows that the procedure is sharp for $N=3$.

Proposition 8 *Our local moment selection procedure provides a sharp characterization of the identified set for $N=3$.*

However, it is difficult to prove sharpness with any number of players N due to the difficulty of globally characterizing all the facets. We evaluate this procedure for $N=4$ in the Monte Carlo section and results highlight that we are sharp too.

4 Estimation and inference

Following the results derived above, we now adopt the approach developed in Beresteanu and Molinari (2008) and Bontemps et al. (2012) for testing a point in a convex set:

$$\begin{aligned}\theta \in \Theta_I(\mathbb{P}) &\Longleftrightarrow P_0 \in A(\theta)\\ &\Longleftrightarrow \forall q \in \mathcal{G},\ T_\infty(q;\theta) = \delta^*(q;A(\theta)) - q^\top P_0 \geq 0\\ &\Longleftrightarrow \min_{q\in\mathcal{G}} T_\infty(q;\theta) \geq 0.\end{aligned}$$

P_0 is the true choice probability. The set of directions $\mathcal{G}$ is defined as

$$\mathcal{G} = \bigcup_{K=0}^{N} \left\{ \mathrm{vec}\Big(0_{\sum_{i=0}^{K-1} d_i}, q_K, 0_{\sum_{i=K+1}^{N} d_i}\Big) : q_K \in \mathcal{G}_K \right\},$$

[16] Intuitively, there are more facets at these extreme points because, due to the lower number of subsets in multiplicity, different orders of the indices $i_1,\ldots,i_{d_K}$ lead to the same point $E^{(K)}_{i_1,\ldots,i_{d_K}}(\theta)$.

where, either $\mathcal{G}_K = \mathcal{Q}_K$ as defined in Proposition 6 or $\mathcal{G}_K = \Omega_K$ the core determining class characterized in Proposition 7.[17] The set $\mathcal{G}$ collects successively all the directions needed to sharply characterize the identified set.

The test based on $T_\infty(\cdot)$ is infeasible because we do not observe P_0. We now characterize the feasible test statistic and its asymptotic distribution under the null and derive strategies to calculate the critical values. Throughout this section, we assume that we observe a sample of M i.i.d. markets in which the same N firms (known to the econometrician) compete.

4.1 The asymptotic distribution of the test statistic

Let $T_M(q;\theta)$ be the empirical counterpart of $T_\infty(q;\theta)$:

$$T_M(q;\theta) = \delta^*(q; A(\theta)) - q^\top \hat{P}_M,$$

where $\hat{P}_M = \dfrac{1}{M}\displaystyle\sum_{m=1}^{M} [\mathbf{1}(Y_m = y_1), \ldots, \mathbf{1}(Y_m = y_{2^N})]^\top$ is the empirical frequency vector. Under the assumption that the markets are i.i.d., we have:

$$\sqrt{M}\left(\hat{P}_M - P_0\right) \xrightarrow[M\to\infty]{d} \mathcal{N}(0, \Sigma_0),$$

where $\Sigma_0 = \mathrm{diag}(P_0) - P_0(P_0)^\top$. Note that the only random part in $T_M(\cdot)$ comes from the estimation of the choice probabilities. Consequently, for q and θ fixed, $T_M(q;\theta)$ is asymptotically normal with variance $q^\top \Sigma_0 q$. An empirical estimator, $\hat{\Sigma}$, can be used by plugging in $\hat{P}_M$ in place of P_0 in the expression of Σ_0.

In this section, we want our asymptotic result to be valid not only for the true probability but also uniformly in the neighborhood of the true probability. We impose the following uniform integrability condition:

Assumption 1 (Uniform integrability) *The class $\mathcal{P}$ satisfies*

$$\lim_{\lambda\to\infty} \sup_{P\in\mathcal{P}} \sup_{j\in\{1,\ldots,2^N\}} \mathbb{E}_P\left[\left(\frac{\mathbf{1}(Y=y_j) - \mu_j(P)}{\sqrt{\mu_j(P)(1-\mu_j(P))}}\right)^2 \mathbf{1}\left\{\left|\frac{\mathbf{1}(y_j) - \mu_j(P)}{\sqrt{\mu_j(P)(1-\mu_j(P))}}\right| > \lambda\right\}\right] = 0, \qquad \text{(UI)}$$

[17] Observe that, if, for any K, $\tilde{q}_K = \mathrm{vec}\Big(0_{\sum_{i=0}^{K-1} d_i}, q_K, 0_{\sum_{i=K+1}^{N} d_i}\Big)$,

$$\delta(\tilde{q}_K; A(\theta)) = \delta(q_K; B_K(\theta)).$$

where $\mu_j(P) = \mathbb{E}_P(Y = y_j)$.

Assumption UI ensures the uniform convergence of our test statistic over the class of probability distributions $\mathcal{P}$. This condition is satisfied over the class of probability distributions such that $\mu_j(P) \geq \varepsilon$ for each j and some $\varepsilon > 0$.

The literature on inference in partially identified models has largely focused on the construction of confidence regions $\mathcal{C}_M$. There is an open debate about whether one should cover any point of the identified set (and, in particular, the true value) or the identified set entirely. Unless stated otherwise, we are mainly interested in covering any point in the identified set with some pre-specified probability $1 - \alpha$, i.e.,

$$\liminf_{M\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P\left(\theta \in \mathcal{C}_M\right) \geq 1 - \alpha.$$

Following Bontemps et al. (2012), our inference method is based on $T_M(q;\theta)$, rescaled by $\sqrt{M}$ and normalized (see Chernozhukov et al., 2015):

$$\xi_M(\theta) = \sqrt{M} \min_{q\in\mathcal{G}} \frac{T_M(q;\theta)}{\sqrt{q^\top \hat{\Sigma} q}}.$$

A point θ belongs to the confidence region $\mathcal{C}$ if the test based on $\xi_M(\theta)$ is not rejected. We now calculate the asymptotic distribution of the test statistics $\xi_M(\theta)$.

Proposition 9 *Let Q_θ be the set of minimizers of $T_\infty(q;\theta)$ in $\mathcal{G}$. Let Z be a random vector of $\mathbb{R}^{2^N}$ distributed according to the normal distribution with variance Σ_0. We have*

$$\begin{cases} \xi_M(\theta) \xrightarrow[M\to\infty]{d} \min_{q\in Q_\theta} \frac{q^\top Z}{\sqrt{q^\top \Sigma q}} & \text{if } P_0 \in A(\theta), \\ \xi_M(\theta) \xrightarrow[M\to\infty]{a.s} -\infty & \text{if } P_0 \notin A(\theta), \end{cases}$$

Under assumption UI, these results are uniformly valid over $P \in \mathcal{P}$.

Observe that the asymptotic distribution depends on θ only through Q_θ. It is an important remark that we exploit in the next subsections to provide new critical values.

4.2 A global bound approach

We first propose conservative bounds. They have the advantage of being calculated simply and once for all for any θ, thereby considerably simplifying the inference procedure. They can be used in a first step to locate the confidence region. This approach is based on the idea that we can trivially define two outer sets of Q_θ and, using these, bound the asymptotic distribution of $\xi_M(\theta)$. We thus have:

$$Q_\theta \subset \mathcal{G} \subset \mathcal{R},$$

where $\mathcal{R}$ is the set of all non null directions of $\mathbb{R}^{2^N}$. Consequently,

$$\min_{q \in Q_\theta} \frac{q^\top Z}{\sqrt{q^\top \Sigma q}} \geq \min_{q \in \mathcal{G}} \frac{q^\top Z}{\sqrt{q^\top \Sigma q}} \geq \min_{q \in \mathcal{R}} \frac{q^\top Z}{\sqrt{q^\top \Sigma q}}.$$

We denote respectively by $c(\mathcal{G}, \alpha)$ and $c(\mathcal{R}, \alpha)$, the α quantile of the bounding distributions. First, observe that

$$\min_{q \in \mathcal{R}} \frac{q^\top Z}{\sqrt{q^\top \Sigma q}} = -\min \|Z^*\|,$$

where Z^* is distributed as a standard multivariate normal variable of dimension 2^N. Second, we provide an algorithm in Section A.15 of the Supplemental Appendix to simulate the critical value $c(\mathcal{G}, \alpha)$ with a number of calculations polynomial in N. The confidence region is defined as the collection of points for which the test statistic $\xi_M(\theta)$ is above a given threshold:

$$\mathcal{C}(c) = \left\{\theta \in \Theta : \quad \xi_M(\theta) \geq c\right\} \tag{11}$$

Using the result of Proposition 9, we can show that the confidence regions built using our procedures have, asymptotically and uniformly over $\mathcal{P}$, the correct coverage rate for any point of the identified set.

Proposition 10 *Under Assumption UI, the following holds:*

$$\liminf_{M \to \infty} \inf_{P \in \mathcal{P}} \inf_{\theta \in \Theta_I} P\left(\theta \in \mathcal{C}(c(\mathcal{R}, \alpha))\right) \geq \liminf_{M \to \infty} \inf_{P \in \mathcal{P}} \inf_{\theta \in \Theta_I} P\left(\theta \in \mathcal{C}(c(\mathcal{G}, \alpha))\right) \geq 1 - \alpha.$$

Observe that because we bound the asymptotic distribution of both $\xi_M(\theta)$ uniformly for $\theta \in \Theta_I$, the confidence regions built from our global bounding strategy entirely cover the identified set. We therefore have the following stronger result:

Corollary 11 *Under Assumption UI,*

$$\liminf_{M \to \infty} \inf_{P \in \mathcal{P}} P\left(\Theta_I \subset \mathcal{C}(c(\mathcal{R}, \alpha))\right) \geq \liminf_{M \to \infty} \inf_{P \in \mathcal{P}} P\left(\Theta_I \subset \mathcal{C}(c(\mathcal{G}, \alpha))\right) \geq 1 - \alpha.$$

4.3 A local bound approach

The global approach above has a tremendous computational advantage because we calculate a critical value once. However, it is quite conservative. We now consider a new bound for the distribution of $\xi_M(\theta)$ based on a moderate deviation inequality for a self normalized sum used in Chernozhukov et al. (2014). It is based on the calculation of an upper bound on the number of facets of each set $B_K(\theta)$, which gives the highest possible number of binding moments.

4.3.1 The number of facets of $B_K(\theta)$ at any extreme point

We know from Proposition 4, that $B_K(\theta)$ is included in an hyperplane of $\mathbb{R}^{d_K}$. It is due to the fact that the sum of the choice probabilities of all outcomes in $\mathcal{Y}_K$ is constant. Following the convex literature, an exposed face is the intersection between a supporting hyperplane and the convex set. Henceforth, we call facets of $B_K(\theta)$ at any extreme point all $d_K - 1$ and $d_K - 2$-faces containing this extreme point. Each facet is related to one supporting hyperplane which defines one irredundant inequality.

From Proposition 4, the maximum number of extreme points of $B_K(\theta)$ is $d_K!$. We first consider the case in which $K = 1$ as a benchmark (by symmetry, it also characterizes the case in which $K = N - 1$) before considering the general case. Observe first that the geometry of a set $B_K(\theta)$ is the same as that of the set $B_{N-K}(\theta)$. We characterize the geometry of the sets $B_1(\theta)$ that will serve as a benchmark for considering the other sets.

Proposition 12 *The convex set $B_1(\theta)$ has $d_1!$ extreme points. Each of them is the intersection of d_1 supporting hyperplanes.*

See the proof in the Supplemental Appendix. Following Proposition 2, we know that any subset of $\mathcal{Y}_1$ of cardinality greater than 2 corresponds to a multiple equilibria region. Consequently, $\Delta_S^{(1)}(\theta)$ for any subset $S \subseteq \mathcal{Y}_1$ is non-zero, and, following Proposition 4, any change in the order gives a different point $E^{(1)}_{i_1,i_2,\ldots,i_{d_1}}$.

For a general value of K, many collections of outcomes cannot be a prediction of a multiple equilibria region. Consequently, there are fewer than $d_K!$ extreme points. Intuitively, different orders of the components of the direction q_K may lead to the same physical extreme point (i.e., for example, $E^{(K)}_{i_1,i_2,i_3\ldots,i_{d_K}} = E^{(K)}_{i_1,i_3,i_2\ldots,i_{d_K}}$). More importantly, there are more than d_K facets at these extreme points. The question is now to determine what are the facets at these specific extreme points and how many they are.

Proposition 13 *Any extreme point of $B_K(\theta)$ for $1 < K < N-1$ is the intersection of at most $2^{l_{\max}} + (d_K - l_{\max} - 1)2^{l_{\max}-1}$ supporting hyperplanes, where $l_{\max}$ is the maximum cardinality of a subset of $\left\{y^{(K)}_1, \ldots, y^{(K)}_{d_K}\right\}$ that cannot be in multiplicity. Furthermore, an algorithm is provided in the Supplemental Appendix to determine the facets at each extreme point $E^{(K)}_{i_1,i_2,i_3\ldots,i_{d_K}}$.*

The result of Proposition 13 gives an upper bound on the number of facets at each extreme point and, therefore, on the number of binding moments. Observe that this number is exponentially smaller than the total number of inequalities $2^{d_K} - 1$. The upper bound on the number of facets passing through any extreme point of $B_K(\theta)$ can be improved further on a case-by-case basis. Table II gives the maximum number of facets for $N = 3$ to 6.

4.3.2 Using the maximum number of facets to calculate a new critical value

The maximum number of facets of $A(\theta)$ at any extreme point is the sum of the maximum number of facets of each $B_K(\theta)$ calculated above. We call it $\mathcal{L}^*$.

Following Chernozhukov et al. (2014), we can use this number to provide a new critical value which is still conservative but much better than the ones derived in the global approach. Intuitively, the critical value calculated in the global approach assumes that all inequalities calculated from the direction in $\mathcal{Q}$ are binding whereas this new one only takes one of the existing extreme points as the worst case. The critical value is equal to

$$c_{\mathcal{L}^*}(\alpha) = \frac{\Phi^{-1}(\alpha/\mathcal{L}^*)}{\sqrt{1 - \Phi^{-1}(\alpha/\mathcal{L}^*)^2/n}} \tag{12}$$

where Φ is the c.d.f. of the standard normal distribution, and Φ^{-1} its inverse. We have now the following result:

Proposition 14 *Under Assumption UI,*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \inf_{\theta\in\Theta_I(P)} P\left(\theta \in \mathcal{C}(c_{\mathcal{L}^*}(\alpha))\right) \geq 1 - \alpha.$$

Observe that the last procedure is not valid for covering the entire identified set because the set of minimizers Q_θ, despite uniformly bounding its cardinal, varies with the point tested.

5 Monte Carlo simulations

We now evaluate the performance of the different testing procedures proposed in Section 4 in a simple game with N players, with N being equal to 3 and 4. The profit function is equal to:

$$\pi_{im} = \beta + \alpha_i \sum_{j \neq i} a_{jm} + \varepsilon_{im},$$

where we assume that the term β is the same across firms, and this is known to the econometrician. ε_{im}, for $i = 1, \ldots, N$, is drawn from a standard Normal distribution. We report the results for $m = 1000$ independent markets. All the results displayed are based on 1000 replications, and the level of the tests is $\alpha = 5\%$.

5.1 Experiments in a game with three players

We consider a model with three symmetric players where $\beta = 0.35$ and $\alpha_1 = \alpha_2 = \alpha_3 = -0.4$. However, the econometrician only knows that β is the same for all players. Consequently, the model is set-identified. The true selection mechanism gives an order of entry to firm 1, 2, 3 in this order in the multiple equilibria regions. Thus, we have the following probabilities for the number of active firms in equilibrium: $P(K = (0, 1, 2, 3)) = (.048, .482, .435, .035)$. We also assume that the profit shocks are independent. For inference, we first compare three different approaches: "Bound" means that we only use the minimum and maximum of the probability of each possible outcome, as in (6), "Ineq" means that we test any point θ by considering the full set of directions proposed in Proposition 6. "Local" means that we apply our geometric selection procedure and only test the inequalities that define this local extreme point.

Critical values We display the results for the choice of two different critical values, $c(\mathcal{G}, \alpha)$ and $c_{\mathcal{L}^*}(\alpha)$ defined in subsections 4.2 and 4.3 (labeled "G" and "L" in the table). For the local approach, we also compare our results with the exact critical value which can be computed, in this case, due to the low dimensionality of the problem.

Before turning to the results, we would like to compare our critical values with the ones which would have been computed while using the moment inequality literature. We consider the one of Andrews and Soares (2010) and its refinement by Romano et al. (2012).

First, it would lead to an increasing amount of computer time. This essentially comes from the fact that the specific structure of the asymptotic distribution is not exploited and that the GMS procedure of Andrews and Soares (2010) has to first evaluate all inequalities, select them and bootstrap the test statistic for each point tested. The computation time drops from 11h to 4s on a single processor (for approximately 2.5 million points tested) when one exploits the geometry and computes the critical values proposed in the last section.

Second, the values are reported in Table III for two points, one on the boundary of the set $A(\theta)$ and one in the interior of the set.[18] For small sample, the GMS procedures do not perform much better than the uniform approach suggested in our paper. One needs a very high sample size, greater than $3,000$ to obtain better performances, and this is only true for the AS procedure for the interior point. Our local approach is, of course, much more accurate for testing a point located at the frontier, it is by construction more conservative for an interior point. Further results about the mean rejection rate across simulations for a sequence of points containing points from the identified set and outside this identified set are provided in the Supplemental Appendix and confirm the fact that our procedure works well.

Simulation results Table IV displays the results of our simulations. We report the information relative to the estimation of α_1, α_2, α_3 and β. We test each point $\theta = (\alpha_1, \alpha_2, \alpha_3, \beta)$ on some predefined grid (defined below each table) and retain those that are not rejected by the testing procedures. The accepted points are then projected on each axis. We report the average of the minimum and maximum values across all simulations. It would be possible to consider the recent work of Kaido et al. (2018), who propose a procedure for subvector inference. However, here, we are mainly interested in comparing the different strategies. We also report the average number of points which pass the test normalized by the average number of points which pass the test in the local approach with the exact critical value. We call it "Vol. CR" in the Tables. This measure is a

[18] They are labeled respectively AS and RSW in the table.

proxy for the relative sizes of the confidence regions.

We also propose different versions of the inequalities. There are many equivalent ways to write inequalities and equalities sufficient to characterize the identified set. The small sample properties of the different testing strategies may be different. This is the reason why we compare the four equivalent sets of inequalities, [19] which all sharply characterize $A(\theta)$. "Ineq$_1$" considers the 16 directions in $\mathcal{Q}$ defined in Proposition 6; in "Ineq$_2$", we replace the directions $(1,1,0)$ in $\mathcal{Q}_K$ for $K=1,2$ by the direction $(0,0,-1)$. It is based on the observation that testing the inequalities for $(1,0,0)^\top$, $(1,1,0)^\top$ and $(1,1,1)^\top$ is equivalent to testing the inequalities for $(1,0,0)^\top$, $(0,0,-1)^\top$ and $(1,1,1)^\top$. "Ineq$_3$" considers the whole set of inequalities and equalities which define $B_0(\theta)$ to $B_3(\theta)$, dropping the equality related to $B_2(\theta)$, because of redundancy. There are 18 inequalities tested. Finally, "Ineq$_4$" applies to Ineq$_3$ the replacement of $(0,0,-1)$ by $(1,1,0)$ in $\mathcal{Q}_K$ for $K=1,2$.

[**Include Tables IV**]

Obviously, a more conservative critical values leads to a larger confidence region. The conservative critical value $c(\mathcal{G},\alpha)$ is not bad at all, even if a local analysis provides a better accuracy. There are some differences between the different sets of inequalities, and the bound approach seems competitive. It is worth noting that the DGP is in favor of the bounds because the independence between the profit shocks make the multiple regions very small.

The most efficient combination is "Ineq$_4$", whatever the choice of the critical value. The geometric selection procedure is conducted with the set of inequalities "Ineq$_4$" and the critical value $c_{\mathcal{L}^*}$. Unsurprisingly, it leads to the same outcome which is very close to the optimal one when one considers the exact critical value, a case difficult to consider when the number of players is higher.

Additional example We also report the results for a slight modification of the parameters ; it changes the accuracy of the bound approach. With $\beta = 0.6$, $\alpha_2 = -0.5$ and $\alpha_1 = \alpha_3 = -0.7$, we have the following probabilities for the number of active firms in equilibrium: $P(K = (0,1,2,3)) = (.021, .499, .464, .016)$. The model is also set-identified and the sizes of $B_1(\theta_0)$ and $B_2(\theta_0)$ are now larger. The results are displayed in Table V. The ratio between the bound approach and the

[19]The directions considered in these four cases are defined in subsection B.3 in the Supplemental Appendix.

"sharp" inequality approach is more in favor of the sharp characterization. "Ineq_4" and our local procedure are still the two best procedures.

[Include Tables V]

5.2 A four player example

We finally consider a four player example. We conduct a similar experiment than for the three player case with two different values of α_i, $i = 1, ..., 4$. We have $\beta = 0.38$, $\alpha_1 = \alpha_4 = -0.35$ and $\alpha_2 = \alpha_3 = -0.2$. The order of entry of the firms in the multiple equilibria regions is purely random. Thus, we have the following probabilities for the number of active firms in equilibrium: $P(K = (0, 1, 2, 3, 4)) = (.015, .237, .530, .207, .011)$. We assume that the econometrician knows that the true values of α are equal two by two. The model is therefore point-identified. The different set of directions considered are defined similarly as well as the critical values. We also report the results relative to our geometric selection procedure ("Local") in which we use $c_{\mathcal{L}^*}$ as the critical value. All the results are displayed in Table VI. First, the size of the confidence region calculated with the bound approach is much larger than the size of the same region calculated with our procedure. When the dimension of $\mathcal{Y}$ increases, the ratio between the volume of the cube which contains $A(\theta)$ and the volume of $A(\theta)$ increases too. A component wise bound approach selects much more points than a local approach (the volume ratio is close to 2). Like, for $N = 3$, the same set of inequalities can be tested differently. Combining equalities and inequalities seem to be better (the volume ratio with respect to the most efficient combination are respectively equal to 1.5 for "Ineq_1" and 1.3 for "Ineq_3"). And, finally, following the results developed in the appendix, a brute force algorithm would test 92 inequalities. A geometric selection procedure leads to the estimation of only 18 inequalities. It is still possible to compare this procedure with the brute force one for this value of N. Observe that they lead to the same size of the confidence region, highlighting the efficiency of our geometric selection procedure.

[Include Table VI]

6 Extensions with explanatory variables

We now generalize our approach to the case with discrete explanatory variables $Z \in \mathbb{R}^l$, whether Z is genuinely discrete or discretized as in Ciliberto and Tamer (2009). Let the support of Z be $\mathcal{Z} = \{z_1, \ldots, z_d\}$. The profit function in Equation (1) becomes

$$\begin{aligned} \pi_{im} &= \beta_i + Z_{im}^\top \gamma_i + \alpha_i \Big(\sum_{j \neq i} a_{jm} \Big) + \varepsilon_{im}, \\ a_{im} &= \mathbf{1}\{\pi_{im} > 0\}. \end{aligned} \tag{13}$$

All the discussion in the previous sections can now be generalized conditional on $Z_m = (Z_{1m}, \ldots, Z_{Nm}) \in \mathcal{Z}^N$. For any realization z_m of Z_m, the set $A_{z_m}(\theta)$ has exactly the same geometry as that described above. Let $P_0(z_m)$ be the true conditional choice probability vector. We obtain the following characterization of the identified set (θ represents all α, β, γ for all firms $i = 1, \ldots, N$):

$$\begin{aligned} \theta \in \Theta_I &\Longleftrightarrow \forall z_m \in \mathcal{Z}^N,\ P_0(z_m) \in A_{z_m}(\theta) \\ &\Longleftrightarrow \forall z_m \in \mathcal{Z}^N,\ \forall q \in \mathcal{G},\ T_\infty(q, z_m; \theta) := \delta^*(q; A_{z_m}(\theta)) - q^\top P_0(z_m) \geq 0 \\ &\Longleftrightarrow \min_{z_m \in \mathcal{Z}^N} \min_{q \in \mathcal{G}} T_\infty(q, z_m; \theta) \geq 0 \end{aligned}$$

The only difference is to add $\min_{z_m \in \mathcal{Z}^N}$ to the procedure, i.e., augment the space over which we take the minimum.

For the inference procedure, we first use a conditional frequency estimator to obtain the conditional choice probabilities. For each $y \in \mathcal{Y}$,

$$\hat{P}_M(y|z_m) = \frac{\sum_{m=1}^M 1[Y_m = y] 1[Z_m = z_m]}{\sum_{m=1}^M 1[Z_m = z_m]}.$$

We define the stacked vector as $\hat{P}_M(z_m)$. This estimator is uniformly consistent, asymptotically normal with a standard convergence speed.

The framework with conditioning variables is nearly the same as before. The only difference comes from the fact that we have to minimize the quantity $\inf_{q \in \mathcal{G}} T(q, z_m; \theta)$ over the discrete space. The simulations of the critical values can be obtained equivalently.

7 Conclusion

In this paper, we develop a new methodology to estimate games with multiple equilibria. The model may be set-identified, and belonging to this identified set is equivalent to testing whether the vector of choice probabilities belongs to a convex set. We characterize the full geometric structure of this convex set without adding any restriction on the selection mechanism in the regions with multiple equilibria. This approach has two advantages. First, it allows us to characterize all the moment inequalities that are necessarily satisfied for a parameter value to be in the identified set. Second, as the complexity of the problem grows exponentially with the number of players, we are able to propose an algorithm that geometrically selects the locally relevant moments, without having to evaluate all of them. The algorithm sequentially approaches the nearest vertex of the polytope, and it requires only the computation of the support function for a number of directions that is polynomial in N, the number of players.

This is a huge improvement in computational burden with respect to alternative methods, which need to numerically evaluate all the moments before deciding which ones are binding and which ones are not. This geometric understanding can easily be extended to other notions of equilibrium proposed in literature, e.g., two-level rationality and social interaction games. We also propose inference methods that for small and moderate sample sizes have better statistical properties than existing methods and offer a considerable advantage in terms of computational burden. Our methods exploit the specific structure of the test statistic, the variance of which does not depend on the parameter tested.

Many pending questions remain. First, entry games generally impose strong restrictions on the functional form of the profit function and on the distribution of the error terms. Adapting our procedure to more general settings (greater heterogeneity or semi-parametric forms) is high on the research agenda. Second, it would be worth investigating inference on subvectors, a question, that has recently been addressed in the partial identification literature, as in Kaido et al. (2018). Finally, we have thus far been agnostic about the (unobserved) selection mechanism, and we have used natural bounds. It would be worth investigating how restrictions on the selection mechanism could be incorporated into this framework.

Acknowledgements

We gratefully acknowledge financial support from the Agence Nationale de Recherche ANR-05-BLAN-0024-01. This research has also received financial support from the European Research Council under the European Community Seventh Framework Program FP7/2007-2013 grant agreement N_295298. We would like to thank the editor, the associate editor and two anonymous referees for helpful comments and participants at various conferences and seminars for helpful discussions.

Bibliography

Andrews, D., and G. Soares (2010). Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection. *Econometrica* **78** 119-157.

Andrews, D., and X. Shi (2013). Inference Based on Conditional Moment Inequalities. *Econometrica* **81** 609-666.

Aradillas-Lopez, A., and A. Rosen (2017). Inference in Ordered Response Games with Complete Information. Working paper.

Aradillas-Lopez, A., and E. Tamer (2008). The Identification Power of Equilibrium in Simple Games. *J. Bus. Econ. Stat.* **26(3)** 261-283.

Bajari, P., H. Hong, and S. Ryan (2010). Identification and Estimation of Discrete Games of Complete Information. *Econometrica* **78** 1529-68.

Beresteanu, A., and F. Molinari (2008). Asymptotic Properties for a Class of Partially Identified Models. *Econometrica* **76(4)** 763-814.

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

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

Berry, S.,and P. Reiss (2007). Empirical Models of Entry and Market Structure, In: M. Armstrong and R. Porter, Editor(s), Handbook of Industrial Organization, Elsevier, 2007, Volume 3.

Berry, S., and E. Tamer (2007). Identification in Models of Oligopoly Entry, in Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, vol. 2, R. Blundell, W.K. Newey and T. Persson, eds., Cambridge Univ. Press.

Bjorn, P., and Q. Vuong (1984). Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation. Working Paper, California Institute of Technology.

Bontemps, C. and T. Magnac (2017). Set Identification, Moment Restrictions and Inference. *Annu. Rev. Econ.* forthcoming.

Bontemps, C., T. Magnac and E. Maurin (2012). Set Identified Linear Models. *Econometrica* **80** 1129-1155.

Bresnahan, T., and P. Reiss (1991a). Entry and Competition in Concentrated Markets. *Journal of Political Economy* **99(5)** 977-1009.

Bresnahan, T., and P. Reiss (1991b). Empirical models of discrete games. *J. Econometrics* **48(1-2)** 57-81.

Bugni, F. (2010). Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities. *Econometrica* **78** 735-754.

Canay, I. (2010). EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity. *J. Econometrics* **156** 408-425.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014) Testing many moment inequalities. arXiv:1312.7614

Chernozhukov, V., H. Hong, and E. Tamer (2007). Inference on Parameter Sets in Econometric Models. *Econometrica* **75** 1243-1284.

Chernozhukov, V., E. Kocatulum and K. Menzel (2015). Inferences on sets in finance. *Quantitative Economics* **6** 309-358.

Chesher A. and A. Rosen (2017). Generalized Instrumental Variable Model. *Econometrica* **85** 959-989.

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

Cleeren, K., M. Dekimpe, K. Gielens, and F. Verboven (2010). Intra- and Inter-format Competition among Discounters and Supermarkets. *Marketing Science* **29(3)** 456-473.

Galichon, A., and M. Henry (2011). Set Identification in Models with Multiple Equilibria. *Rev. Econ. Stud.* **78** 1264-1298.

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

Gualdini C. (2018). An Econometric Model of Network Formation with an Application to Board Interlocks between Firms. TSE Working Paper 17-898, Toulouse School of Economics.

Kaido, H., F. Molinari, and J. Stoye (2018). Confidence Intervals for Projections of Partially Identified Parameters. Working paper.

Kaido, H., and A. Santos (2014). Asymptotically efficient estimation of models defined by convex moment inequalities. *Econometrica* **82(1)** 387-413.

Khan S. and E. Tamer (2010). Irregular Identification, Support Conditions, and Inverse Weight Estimation. *Econometrica* **78(6)** 2021–2042.

Mazzeo M. (2002). Product Choice and Oligopoly Market Structure. *Rand Journal of Economic* **33(2)** 1-22.

Menzel, K. (2014). Inference for Games with Many Players. *Rev. Econ. Stud.* **83(1)** 306-337.

Molchanov, I., and F. Molinari (2015). Applications of random set theory in econometrics. *Annu. Rev. Econ.* **6(1)** 229-251.

de Paula A. (2013). Econometric Analysis of Games with Multiple Equilibria. *Annu. Rev. Econ.* **5** 107-131.

Reiss, P. (1996). Empirical Models of Discrete Strategic Choices. *American Economic Review* **86** 421-426.

Rockafellar, R.T. (1970). *Convex Analysis.* Princeton University Press.

Romano, J.P., and A.M. Shaikh (2008). Inference for Identifiable Parameters in Partially Identified Econometric Models. *J. Statist. Plann. Infer.* **138** 2786-2807.

Romano, J.P., and A.M. Shaikh (2010). Inference for the identified set in partially identified econometric models. *Econometrica* **78(1)** 169-211.

Romano, J.P., and A.M. Shaikh (2012). On the Uniform Asymptotic Validity of Subsampling and the Bootstrap. *Ann. Stat.*, **40(6)** 2798-2822.

Romano, J., A. Shaikh, and M. Wolf (2014). A Practical Two-Step Method for Testing Moment Inequalities. *Econometrica* **82(5)** 1979-2002.

Rosen, A. (2008). Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. *J. Econometrics* **146** 107-117.

Sampaio R. (2007). Competition Between Low-Cost Carriers and Traditional Airlines: An Empirical Entry Model. TSE *working paper.*

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *Rev. Econ. Stud.* **70(1)** 147-165.

A Figures and Tables

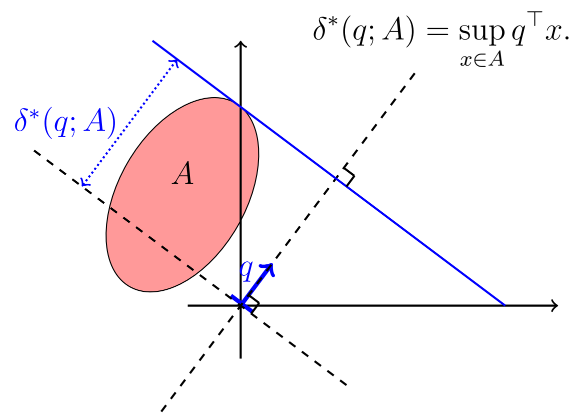


Figure 1: The support function

N	K	d_K	$\lvert S^{(K)} \rvert$	$2^{d_K} - d_K - 1$
3	1	3	4	4
	2	3	4	4
4	1	4	11	11
	2	6	21	59
	3	4	11	11
5	1	5	26	26
	2	10	71	1018
	3	10	71	1018
	4	5	26	26
6	1	6	57	57
	2	15	198	32761
	3	20	283	1048569
	4	15	198	32761
	5	6	57	57

Table I: Counting the number of multiple equilibria regions

	$B_1(\theta)$	$B_2(\theta)$	$B_3(\theta)$	$B_4(\theta)$	$B_5(\theta)$
$N = 4$	4	10	4		
$N = 5$	5	18	18	5	
$N = 6$	6	52	136	52	6

Table II: Upper bound on the number of facets at any extreme point for $N = 4, 5, 6$.

	Extreme point					Interior point				
	True	G	L	AS	RSW	True	G	L	AS	RSW
500	-2.555	-2.747	-2.562	-2.695	-2.771	-2.386	-2.747	-2.562	-2.715	-2.778
1000	-2.555	-2.747	-2.562	-2.655	-2.738	-2.386	-2.747	-2.562	-2.690	-2.774
2000	-2.555	-2.747	-2.562	-2.629	-2.697	-2.386	-2.747	-2.562	-2.630	-2.743
3000	-2.555	-2.747	-2.562	-2.617	-2.677	-2.386	-2.747	-2.562	-2.555	-2.705
10000	-2.555	-2.747	-2.562	-2.581	-2.634	-2.386	-2.747	-2.562	-2.394	-2.455

We compare the critical values of different procedures with the DGP of the set-identified model with 3 players, for two particular points tested. One point is an extreme point, i.e. 8 inequalities are binding in this point. One is an "interior" point, i.e. no inequality is binding but the ones related to the number of players.

Table III: Critical values for different procedures.

Type	Test	Crit. value	α_1 min	α_1 max	α_2 min	α_2 max	α_3 min	α_3 max	β min	β max	Vol. CR
	Bound	G	-0.657	-0.166	-0.715	-0.227	-0.684	-0.197	0.253	0.492	1.49
		L	-0.631	-0.184	-0.688	-0.243	-0.658	-0.214	0.263	0.479	1.06
	Ineq_1	G	-0.714	-0.118	-0.755	-0.165	-0.743	-0.152	0.120	0.488	2.02
		L	-0.693	-0.134	-0.731	-0.180	-0.721	-0.167	0.132	0.478	1.59
	Ineq_2	G	-0.655	-0.094	-0.713	-0.155	-0.683	-0.121	0.109	0.491	1.76
		L	-0.630	-0.114	-0.688	-0.175	-0.658	-0.141	0.126	0.478	1.29
	Ineq_3	G	-0.721	-0.176	-0.762	-0.232	-0.750	-0.208	0.255	0.491	1.77
		L	-0.704	-0.185	-0.742	-0.241	-0.732	-0.217	0.258	0.483	1.45
	Ineq_4	G	-0.655	-0.165	-0.716	-0.228	-0.682	-0.196	0.252	0.495	1.36
		L	-0.634	-0.179	-0.695	-0.240	-0.661	-0.209	0.258	0.483	1.05
	Local	L	-0.634	-0.179	-0.695	-0.240	-0.661	-0.209	0.258	0.483	1.05
		Exact	-0.630	-0.181	-0.692	-0.242	-0.658	-0.212	0.259	0.481	1.00

Note: the test statistic is $\xi_M(\theta)$ calculated on a grid of points $(\alpha_1, \alpha_2, \alpha_3, \beta)$: $[-1.5; 0[$ for the alpha's with a tick of 0.03 and $[0; 1.2]$ for β with a tick of 0.02. The level is equal to 5%. We display the mean across one thousand simulations of the minimum and maximum of each parameter. "Vol. CR" is the mean volume of the confidence regions normalized by the mean volume of the confidence regions computed with the local procedure with the exact critical value. The true values are $\beta = 0.35$ and $\alpha_1 = \alpha_2 = \alpha_3 = -0.4$, the selection mechanism gives priority to firm 1 than to firm 2 to enter in the multiple equilibria regions; the econometrician does not assume that the $\alpha's$ are the same. The model is set-identified.

Table IV: The confidence region - 3 players, set-identified case, 1000 markets.

Type	Test	Crit. value	α_1 min	α_1 max	α_2 min	α_2 max	α_3 min	α_3 max	β min	β max	Vol. CR
	Bound	G	-1.146	-0.429	-0.973	-0.233	-1.150	-0.434	0.484	0.843	1.92
		L	-1.097	-0.446	-0.922	-0.250	-1.101	-0.450	0.492	0.809	1.36
	Ineq_1	G	-1.156	-0.299	-0.986	-0.115	-1.160	-0.302	0.243	0.823	1.85
		L	-1.124	-0.315	-0.954	-0.130	-1.129	-0.318	0.257	0.803	1.49
	Ineq_2	G	-1.137	-0.281	-0.955	-0.102	-1.141	-0.285	0.231	0.835	1.75
		L	-1.092	-0.301	-0.909	-0.122	-1.096	-0.305	0.252	0.803	1.30
	Ineq_3	G	-1.172	-0.440	-1.002	-0.253	-1.175	-0.444	0.487	0.835	1.58
		L	-1.143	-0.449	-0.973	-0.261	-1.147	-0.453	0.489	0.816	1.31
	Ineq_4	G	-1.133	-0.431	-0.943	-0.238	-1.137	-0.435	0.483	0.838	1.33
		L	-1.099	-0.443	-0.908	-0.252	-1.102	-0.448	0.489	0.816	1.04
	Local	L	-1.099	-0.443	-0.908	-0.252	-1.102	-0.448	0.489	0.816	1.04
		Exact	-1.093	-0.446	-0.904	-0.254	-1.097	-0.451	0.491	0.814	1.00

Note: the test statistic is $\xi_M(\theta)$ calculated on a grid of points $(\alpha_1, \alpha_2, \alpha_3, \beta)$: $[-1.5; 0[$ for the alpha's with a tick of 0.03 and $[0; 1.2]$ for β with a tick of 0.02. The level is equal to 5%. We display the mean across one thousand simulations of the minimum and maximum of each parameter. "Vol. CR" is the mean volume of the confidence regions normalized by the mean volume of the confidence regions computed with the local procedure. The true values are $\beta = 0.6$, $\alpha_2 = -0.5$ and $\alpha_1 = \alpha_3 = -0.7$ the selection mechanism gives priority to firm 1 than to firm 2 to enter in the multiple equilibria regions; the econometrician does not assume that α_1 and α_3 are the same. The model is set-identified.

Table V: The confidence region - 3 players, set-identified case, 1000 markets.

Test	Crit. value	α_1 min	α_1 max	α_2 min	α_2 max	β min	β max	Vol. CR
Bound	G	-0.602	-0.232	-0.455	-0.062	0.253	0.682	2.60
	L	-0.576	-0.240	-0.430	-0.072	0.260	0.648	2.04
Ineq_1	G	-0.573	-0.242	-0.421	-0.095	0.176	0.631	2.21
	L	-0.542	-0.254	-0.394	-0.106	0.197	0.593	1.55
Ineq_3	G	-0.575	-0.247	-0.423	-0.097	0.259	0.633	1.80
	L	-0.547	-0.258	-0.399	-0.108	0.271	0.600	1.30
Ineq_4	G	-0.561	-0.243	-0.417	-0.083	0.250	0.645	1.55
	L	-0.525	-0.259	-0.381	-0.101	0.267	0.598	1.00
Local	L	-0.526	-0.258	-0.381	-0.100	0.267	0.598	1.00

Note: the test statistic is $\xi_M(\theta)$ calculated on a grid of points $(\alpha_1, \alpha_2, \beta)$: $[-1.5; 0[$ for the alpha's and $[0; 1.2]$ for β with a tick of 0.01. The level is equal to 5%. We display the mean across one thousand simulations of the minimum and maximum of each parameter. "Vol. CR" is the mean volume of the confidence regions normalized by the mean volume of the confidence regions computed with the local procedure. The true values are $\beta = 0.38$, $\alpha_1 = \alpha_4 = -0.35$ and $\alpha_2 = \alpha_3 = -0.2$, the order of entry of the firms in the multiple equilibria regions is random. The model is point-identified.

Table VI: The confidence region - 4 players, point-identified case, 1000 markets.

Supplemental Material

Supplement to "A Geometric Approach to Inference in Set-Identified Entry Games"

by Christian Bontemps and Rohit Kumar

A Proofs and algorithms

A.1 Proof of Proposition 1

First, observe that, by a revealed preference argument, the region of ε that corresponds to an outcome $y_j^{(K)}$ with K active firms in equilibrium is included in the region:

$$\mathcal{R}(y_j^{(K)}) = \left\{ \varepsilon = (\varepsilon_1, \dots, \varepsilon_N) : \begin{array}{ll} \varepsilon_i \leq -\beta_i - K\alpha_i & \text{if } y_{j,i}^{(K)} = 0 \\ \varepsilon_i > -\beta_i - (K-1)\alpha_i & \text{if } y_{j,i}^{(K)} = 1 \end{array} \right\}.$$

Without loss of generality, assume $S = \{y_1^{(K)}, \dots, y_m^{(K)}\}$ is a collection of outcomes in multiplicity. We first characterize the region of ε that generates this set of outcomes. First, we define three subsets of $\{1, \dots, N\}$. N_0 is the set of indices for which the action of player i is 0 for all outcomes in S, and N_1 is the set of indices i for which the action of player i is 1 for all outcomes in S. The remaining set N_s corresponds to the players who play actions 0 or 1 across the outcomes of S. Without loss of generality, we assume that $N_0 = \{1, 2, \dots, n_0\}$, $N_1 = \{n_0+1, n_0+2, \dots, n_0+n_1\}$, and $N_s = \{n_0+n_1+1, \dots, N\}$. We now prove that $\mathcal{R}_S^{(K)}(\theta)$, the region of ε that predicts all outcomes in S, is defined as follows:

$$\mathcal{R}_S^{(K)}(\theta) = \left\{ \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix} : \begin{pmatrix} \varepsilon_i \leq -\beta_i - (K-1)\cdot\alpha_i & i \leq n_0 \\ \varepsilon_i > -\beta_i - K\cdot\alpha_i & n_0 < i \leq n_0 + n_1 \\ -\beta_i - (K-1)\cdot\alpha_i < \varepsilon_i \leq -\beta_i - K\cdot\alpha_i & i > n_0 + n_1 \end{pmatrix} \right\}.$$

First, take ε in the region defined by the right hand side. Each firm 1 to n_0 is only profitable when it has $K-2$ competitors, each firm n_0+1 to n_0+n_1 is profitable with K competitors, and each of the remaining firms is profitable with $K-1$ competitors. In a situation with complete information, firms n_0+1 to n_0+n_1 enter the market, firms 1 to n_0 do not enter, and $K-n_1$ firms out of the last $n_s = N - n_0 - n_1$ enter. There are therefore $\binom{N-n_0-n_1}{K-n_1}$ possibilities.

Conversely, consider ε in $\mathcal{R}_S^{(K)}(\theta)$, i.e. assume that there is a region which predicts all the outcomes in S as possible equilibria. Obviously, this region is contained in $\bigcap_{1\leq j\leq m} \mathcal{R}(y_j^{(K)})$. If, for $i \leq n_0$, one of the profit shocks ε_i were between $-\beta_i-(K-1)\cdot\alpha_i$ and $-\beta_i-K\cdot\alpha_i$, the corresponding firms could enter to replace one of the last n_s firms, which is in contradiction with the fact that the model predicts all the outcomes in S only. Thus, in fact, $\varepsilon_i \leq -\beta_i-(K-1)\cdot\alpha_i$, $i\leq n_0$. Similarly, $\varepsilon_i > -\beta_i - K\cdot\alpha_i$ for $n_0 < i \leq n_0+n_1$. This proves the reverse inclusion. The cardinality of S is therefore $\binom{N-n_0-n_1}{K-n_1}$.

A.2 Proof of Proposition 2

Following Proposition 1, any set $S=\{y_1^{(K)},\ldots,y_m^{(K)}\}$ of outcomes in multiplicity is such that there are n_0 firms that never enter, n_1 firms that always enter and $n_s = N-n_0-n_1$ that enter in some outcomes and do not in others, with there being in total $K-n_1$ entering among these n_s firms for each outcome (thus, $n_s > K-n_1$). Obviously, $n_1 \leq K-1$ because S contains at least two different outcomes. There are $\binom{N}{n_1}$ choices for these n_1 firms. Among the remaining $N-n_1$, n_0 never enter and $n_s = N-n_0-n_1$ "switch" across outcomes. For each value of n_0, there are $\binom{N-n_1}{n_0}$ choices for each choice of the n_1 firms. As $n_s \geq K-n_1+1$, n_0 is therefore bounded above by $N-K-1$. The number of multiple equilibria regions is equal to:

$$s_K = \sum_{n_1=0}^{K-1}\sum_{n_0=0}^{N-K-1}\binom{N}{n_1}\binom{N-n_1}{n_0}.$$

A.3 Proof of Proposition 3

The convexity of the set $A(\theta)$ can be easily proved from the expression of the $P_j^{(K)}(\theta,\eta)$'s in Equation (3) for any $K=0,\ldots,N$ and any $j=1,\ldots,d_K$. Let $\lambda\in[0,1]$, $\eta_1(\cdot)$ and $\eta_2(\cdot)$ two selection mechanisms and $P_1 = P(\theta,\eta_1)$ and $P_2=P(\theta,\eta_2)$, two vectors of choice probabilities. First, $\eta(\cdot) = \lambda\eta_1(\cdot)+(1-\lambda)\eta_2(\cdot)$ is also a selection mechanism. Second, for any $K=0,\ldots,N$ and any $j = 1,\ldots,d_K$,

$$P_j^{(K)}(\theta,\eta) = P_j^{(K)}(\theta,\lambda\eta_1+(1-\lambda)\eta_2) = \lambda P_j^{(K)}(\theta,\eta_1)+(1-\lambda)P_j^{(K)}(\theta,\eta_2).$$

Consequently, $P(\theta,\eta)=\lambda P_1+(1-\lambda)P_2$.

For the cartesian product, consider two different ε_1 and ε_2 in $\mathcal{R}_S^{(K_1)}(\theta)$ and $\mathcal{R}_{S'}^{(K_2)}(\theta)$, for $K_1 \neq K_2$; the equilibrium selection mechanism is equal to zero when $y \in \mathcal{Y}_{K_1}$ and $\varepsilon = \varepsilon_2$ or when $y \in \mathcal{Y}_{K_2}$ and $\varepsilon = \varepsilon_1$.

A.4 Proof of Proposition 4

We introduce some useful additional notation. For any $S \in S^{(K)}$ and any $j \leq d_K$, we define

$$u_j(S) = \frac{\int_{\mathcal{R}_S^{(K)}(\theta)} \eta(y_j^{(K)}|\varepsilon, \theta) dF(\varepsilon; \theta)}{\int_{\mathcal{R}_S^{(K)}(\theta)} dF(\varepsilon; \theta)}$$

and set $u_j(S) = 0$ when $S \notin S^{(K)}$. Note that for all j such that $y_j^{(K)} \notin S$, $u_j(S) = 0$ because a ε in $\mathcal{R}_S^{(K)}(\theta)$ does not predict $y_j^{(K)}$ as a potential outcome. By construction, $0 \leq u_j(S) \leq 1$ and

$$\sum_{j|\, y_j^{(K)} \in S} u_j(S) = 1.$$

for any $S \in S^{(K)}$. We also define the possibility set for $u_j(S)$, $j = 1, \ldots, d_K$ as

$$U^{(K)}(S) = \left\{ u_j(S) \in [0;1], j = 1, \ldots, d_K, \text{ such that } \sum_{j|\, y_j^{(K)} \in S} u_j(S) = 1 \text{ and } u_j(S) = 0, \text{ if } y_j^{(K)} \notin S \right\},$$

and

$$\mathcal{U}^{(K)} = \left\{ U^{(K)}(S), S \in S^{(K)} \right\}.$$

Based on this additional notation, we can define the set $B_K(\theta)$ as

$$B_K(\theta) = \left\{ P^{(K)} : P_j^{(K)} = \Delta_j^{(K)}(\theta) + \sum_{S \in S_j^{(K)}} u_j(S) \Delta_S^{(K)}(\theta), \;\; u_j(S) \in U^{(K)}(S), j = 1, \ldots, d_K, S \in S^{(K)} \right\}. \tag{A.1}$$

Following the definition of the support function:

$$\begin{aligned} \delta\left(q_K; B_K(\theta)\right) &= \sup_{P^{(K)} \in B_K(\theta)} q_K^\top P^{(K)} \\ &= \sum_{j=1}^{d_K} q_{j,K} \Delta_j^{(K)}(\theta) + \sup_{u_j(S) \in \mathcal{U}^{(K)}} \sum_{j=1}^{d_K} \left(q_{j,K} \sum_{S \in S_j^{(K)}} u_j(S) \Delta_S^{(K)}(\theta) \right) \end{aligned}$$

$$= \sum_{j=1}^{d_K} q_{j,K}\Delta_j^{(K)}(\theta) + \sup_{u_j(S)\in\mathcal{U}^{(K)}} \sum_{j=1}^{d_K} \left(q_{j,K} \sum_{S\in S^{(K)}} u_j(S)\Delta_S^{(K)}(\theta) \right).$$

The last equality (the sum is indexed by $S \in S^{(K)}$ instead of $S \in S_j^{(K)}$) is true because $u_j(S)$ is equal to zero when $S \notin S_j^{(K)}$.

Consequently,

$$\begin{aligned}
\sup_{u_j(S)\in\mathcal{U}^{(K)}} \sum_{j=1}^{d_K} \left(q_{j,K} \sum_{S\in S^{(K)}} u_j(S)\Delta_S^{(K)}(\theta) \right) &= \sup_{u_j(S)\in\mathcal{U}^{(K)}} \sum_{S\in S^{(K)}} \Delta_S^{(K)}(\theta) \left(\sum_{j=1}^{d_K} q_{j,K} u_j(S) \right) \\
&= \sum_{S\in S^{(K)}} \Delta_S^{(K)}(\theta) \sup_{u_j(\cdot)\in U^{(K)}(S)} \left(\sum_{j=1}^{d_K} q_{j,K} u_j(S) \right) \\
&= \sum_{S\in S^{(K)}} \Delta_S^{(K)}(\theta) \left(\max_{j|y_j^{(K)}\in S} q_{j,K} \right)
\end{aligned}$$

Thus,

$$\delta\left(q; B_K(\theta)\right) = \sum_{j=1}^{d_K} q_{j,K}\Delta_j^{(K)}(\theta) + \sum_{S\in S^{(K)}} \Delta_S^{(K)}(\theta) \left(\max_{j|y_j^{(K)}\in S} q_{j,K} \right)$$

We can therefore reorder according to the new partition $\mathcal{O}_{i_1}^{(K)}, \mathcal{O}_{i_2}^{(K)}, \ldots, \mathcal{O}_{i_{d_K}}^{(K)}$.

$$\delta\left(q; B_K(\theta)\right) = \sum_{j=1}^{d_K} q_{j,K}\Delta_j^{(K)}(\theta) + \sum_{j=1}^{d_K} q_{i_j,K} \left\{ \sum_{S\in\mathcal{O}_{i_j}^{(K)}} \Delta_S^{(K)}(\theta) \right\}.$$

A.5 Proof of Proposition 6

We only need to prove the following result:

$$\forall K \in \{0, 1, 2, \ldots, N\}, \ \ \forall q_K \in \mathcal{Q}_K,\ q_K^\top P_0^{(K)} \le \delta^*(q_K; B_K(\theta)) \Longrightarrow \forall q \in \mathbb{R}^{2^N},\ q^\top P_0 \le \delta^*(q; A(\theta)).$$

Let q_K be a direction of $\mathbb{R}^{d_K}$ and assume that its components are ranked in the following order

$$q_{i_1,K} \ge q_{i_2,K} \ge \ldots \ge q_{i_{d_K},K}.$$

Let the d_K directions of $\mathbb{R}^{d_K}$, $e_{i_1}^{(K)}$, $e_{i_1,i_2}^{(K)}$,..., $e_{i_1,i_2,\ldots,i_{d_K-1}}^{(K)}$, $e_{i_1,i_2,\ldots,i_{d_K}}^{(K)}$, where the components are equal to 1 when the indices are subscripts of $e^{(K)}$ and 0 otherwise. Obviously, these directions belong to $\mathcal{Q}_K$.[20] We can write q_K as a function of these directions with non-negative weights:

$$q_K = (q_{i_1,K} - q_{i_2,K})e^{(K)}_{i_1} + (q_{i_2,K} - q_{i_3,K})e^{(K)}_{i_1,i_2} + \ldots + (q_{i_{d_K-1},K} - q_{i_{d_K},K})e^{(K)}_{i_1,i_2,\ldots,i_{d_K-1}} + q_{i_{d_K},K}e^{(K)}_{1,2,\ldots,d_K}.$$

Assume that the inequalities $\tilde{q}_K^\top P_0^{(K)} \leq \delta^*(\tilde{q}_K; B_K(\theta))$ are satisfied for any direction $\tilde{q}_K \in \mathcal{Q}_K$. We have:

$$\begin{aligned} q_K^\top P_0^{(K)} &= (q_{i_1,K} - q_{i_2,K})(e^{(K)}_{i_1})^\top P_0^{(K)} + (q_{i_2,K} - q_{i_3,K})(e^{(K)}_{i_1,i_2})^\top P_0^{(K)} + \ldots + q_{i_{d_K},K}(e^{(K)}_{1,2,\ldots,d_K})^\top P_0^{(K)} && \text{(A.2)} \\ &\leq (q_{i_1,K} - q_{i_2,K})\delta^*(e^{(K)}_{i_1}; B_K(\theta)) + (q_{i_2,K} - q_{i_3,K})\delta^*(e^{(K)}_{i_1,i_2}; B_K(\theta)) + \ldots + q_{i_{d_K},K}\delta^*(e^{(K)}_{1,2,\ldots,d_K}; B_K(\theta)) && \text{(A.3)} \\ &\leq \delta^*((q_{i_1,K} - q_{i_2,K})e^{(K)}_{i_1}; B_K(\theta)) + \delta^*((q_{i_2,K} - q_{i_3,K})e^{(K)}_{i_1,i_2}; B_K(\theta)) + \ldots + \delta^*(q_{i_{d_K},K}e^{(K)}_{1,2,\ldots,d_K}; B_K(\theta)) && \text{(A.4)} \\ &\leq \delta^*((q_{i_1,K} - q_{i_2,K})e^{(K)}_{i_1} + (q_{i_2,K} - q_{i_3,K})e^{(K)}_{i_1,i_2} + \ldots + q_{i_{d_K},K}e^{(K)}_{1,2,\ldots,d_K}; B_K(\theta)) = \delta^*(q_K; B_K(\theta)) && \text{(A.5)} \end{aligned}$$

Inequality (A.3) comes from the fact that the directions $e^{(K)}_{i_1,\ldots}$ belong to $\mathcal{Q}_K$; inequality (A.4) holds because the support function is positive homogeneous; inequality (A.5) is due to the subadditivity of the support function. Consequently:

$$q^\top P_0 = \sum_{K=0}^{N} q_K^\top P_0^{(K)} \leq \sum_{K=0}^{N} \delta^*(q_K; B_K(\theta)) = \delta^*(q; A(\theta)).$$

A.6 Proof of Proposition 7

Note that $\mathcal{Y}_K$ is well connected (the empty set is connected). Thus, $\mathcal{Y}_K \in \Omega_K$ for $K = 0, 1, \ldots, N$. If the inequality holds for all well-connected subsets, then

$$\mathbb{P}(\mathcal{Y}_K) \leq \mathcal{L}(\mathcal{Y}_K) \tag{A.6}$$

However, we also know that

$$\mathcal{L}(\mathcal{Y}_K \cup \mathcal{Y}_{K'}) = \mathcal{L}(\mathcal{Y}_K) + \mathcal{L}(\mathcal{Y}_{K'})$$

because $\Gamma_{\mathcal{Y}_K}$ is a component of $\Gamma_{\mathcal{Y}}$, i.e., there is no multiplicity between $\mathcal{Y}_K$ and $\mathcal{Y}_{K'}$. As $\mathcal{L}(\mathcal{Y}) = 1$, we have

$$\sum_{K=0}^{N} \mathcal{L}(\mathcal{Y}_K) = 1 \tag{A.7}$$

[20] Observe that $e^{(K)}_{i_1,i_2,\ldots,i_{d_K}} = e^{(K)}_{1,2,\ldots,d_K}$.

From Equations (A.6) and (A.7), we have

$$\begin{aligned}\mathbb{P}(\mathcal{Y}_K) &= 1 - \sum_{i \neq K} \mathbb{P}(\mathcal{Y}_i) \\ &\geq 1 - \sum_{i \neq K} \mathcal{L}(\mathcal{Y}_i) = \mathcal{L}(\mathcal{Y}_K),\end{aligned}$$

where the second line uses Inequality (A.6) and the last equality comes from (A.7). Finally, we have

$$\mathbb{P}(\mathcal{Y}_K) = \mathcal{L}(\mathcal{Y}_K), \tag{A.8}$$

for $K = 0, 1, \ldots, N$.

• We first show that if the inequality holds for all well connected subset, then for any subset $C \subset \mathcal{Y}$, $\mathbb{P}(C) \leq \mathcal{L}(C)$.

Assume that $C = \cup_{K=1}^{N} C_K$, where $C_K \subset \mathcal{Y}_K$. Following Corollary 9, $\mathcal{L}(C) = \sum_{K=1}^{N} \mathcal{L}(C_K)$. Thus, if the inequality holds for each C_K, then it clearly holds for C. Without loss of generality, we now assume that there is one K such that $C \subset \mathcal{Y}_K$.

If C is not well connected, then $\mathcal{Y}_K \setminus C$ is not connected in $\Gamma_{\mathcal{Y}_K \setminus C}$. Therefore, $\mathcal{Y}_K \setminus C$ is a disjoint union of p components $\{W_i\}_{i=1}^{p}$ of the graph $\Gamma_{\mathcal{Y}_K \setminus C}$. Define, for each i in $1, \ldots, p$

$$B_i = C \cup W_1 \cup \ldots \cup W_{i-1} \cup W_{i+1} \cup \ldots \cup W_p$$

$\mathcal{Y}_K \setminus B_i = W_i$ is connected in $\Gamma_{\mathcal{Y}_K \setminus B_i} = \Gamma_{W_i}$ because W_i is a component of the graph $\Gamma_{\mathcal{Y}_K \setminus C}$. B_i is therefore well connected, and thus, $B_i \in \Omega_K$. Therefore, by definition,

$$\mathbb{P}(B_i) \leq \mathcal{L}(B_i).$$

We can now impose a lower bound on $\mathbb{P}(W_i)$ using $\mathbb{P}(W_i) + \mathbb{P}(B_i) = \mathbb{P}(\mathcal{Y}_K)$.

$$\mathbb{P}(W_i) \geq \mathcal{L}(\mathcal{Y}_K) - \mathcal{L}(B_i),$$

because $\mathbb{P}(\mathcal{Y}_K) = \mathcal{L}(\mathcal{Y}_K)$ from Equation (A.8) and the inequality above.

We can now impose an upper bound on $\mathbb{P}(C)$:

$$\mathbb{P}(C) = \mathbb{P}(\mathcal{Y}_K) - \mathbb{P}(\mathcal{Y}_K \setminus C)$$

$$= \mathbb{P}(\mathcal{Y}_K) - \sum_{i=1}^{k} \mathbb{P}(W_i)$$

$$\leq \mathcal{L}(\mathcal{Y}_K) - \sum_{i=1}^{p} \left[\mathcal{L}(\mathcal{Y}_K) - \mathcal{L}(B_i)\right]$$

We finally prove that the last term is $\mathcal{L}(C)$. For each i, following the definition of the Choquet capacity, $\mathcal{L}(\mathcal{Y}_K) - \mathcal{L}(B_i)$ is the sum of probabilities of the unique regions of outcomes of W_i and multiplicity regions only involving outcomes of W_i. Since W_i is not connected to W_j in $\Gamma_{\mathcal{Y}_K \backslash C}$, $\sum_{i=1}^{p} \left[\mathcal{L}(\mathcal{Y}_K) - \mathcal{L}(B_i)\right]$ is the probability of the unique region of outcomes of $\mathcal{Y}_K \backslash C$ and multiplicities only involving outcomes of $\mathcal{Y}_K \backslash C$. Hence, $\mathcal{L}(\mathcal{Y}_K) - \sum_{i=1}^{p} \left[\mathcal{L}(\mathcal{Y}_K) - \mathcal{L}(B_i)\right]$ is the sum of probabilities of unique regions of outcomes in C and multiplicity regions involving only outcomes in C. This is $\mathcal{L}(C)$. We therefore have

$$\mathbb{P}(C) \leq \mathcal{L}(C).$$

• We now prove that if a well-connected subset B of $\mathcal{Y}_K$ is not part of Ω_K, we can define a DGP where all inequalities $\mathbb{P}(C) \leq \mathcal{L}(C)$ hold except in B, thus violating the assumption that Ω is core determining for $\mathcal{L}$.

Assume that there are p elements in B, $y_1, \ldots, y_p$. We omit in the proof the superscript (K) for ease of exposition. For a given $\varepsilon > 0$, we consider the following probability outcome. It is an outcome in which we reallocate some of the predictions in the multiple equilibria regions from the first outcome y_1 and y_{p+1} to the $p-1$ outcomes $y_2, \ldots, y_p$.

$$\begin{aligned}
\mathbb{P}(y_1) &= \mathcal{L}(y_1) - (p-2)\varepsilon \\
\mathbb{P}(y_2) &= \mathcal{L}(\{y_1, y_2\}) - \mathcal{L}(y_1) + \varepsilon \\
&\vdots \\
\mathbb{P}(y_{p-1}) &= \mathcal{L}(\{y_1, \ldots, y_{p-1}\}) - \mathcal{L}(\{y_1, \ldots, y_{p-2}\}) + \varepsilon \\
\mathbb{P}(y_p) &= \mathcal{L}(\{y_1, \ldots, y_p\}) - \mathcal{L}(\{y_1, \ldots, y_{p-1}\}) + \varepsilon \\
&- \\
\mathbb{P}(y_{p+1}) &= \mathcal{L}(\{y_1, \ldots, y_{p+1}\}) - \mathcal{L}(\{y_1, \ldots, y_p\}) - \varepsilon \\
\mathbb{P}(y_{p+2}) &= \mathcal{L}(\{y_1, \ldots, y_{p+2}\}) - \mathcal{L}(\{y_1, \ldots, y_{p+1}\})
\end{aligned}$$

$$\vdots$$

$$\mathbb{P}(y_{|\mathcal{Y}_K|}) = \mathcal{L}(\{y_1, \ldots, y_{|\mathcal{Y}_K|}\}) - \mathcal{L}(\{y_1, \ldots, y_{|\mathcal{Y}_K|-1}\})$$

Our goal is to show that the inequalities $\mathbb{P}(C) \leq \mathcal{L}(C)$ are satisfied for all elements of $\mathcal{Y}_{\mathcal{K}}$ but B for some adequate choice of ε.

First, note that the violation of the inequality for B is obvious because:

$$\begin{aligned}
\mathbb{P}(B) &= \sum_{i=1}^{p} \mathbb{P}(y_i) \\
&= \sum_{i=2}^{p} \Big[\mathcal{L}(\{y_1, \ldots, y_i\}) - \mathcal{L}(\{y_1, \ldots, y_{i-1}\}) + \varepsilon \Big] + \mathcal{L}(y_1) - (p-2)\varepsilon \\
&= \mathcal{L}(\{y_1, \ldots, y_p\}) + \varepsilon \\
&= \mathcal{L}(B) + \varepsilon
\end{aligned}$$

Now, we show that no other inequality is violated for this constructed probability under some condition on ε.

(i) Find r such that $y_2, \ldots, y_r \in B$ are directly connected to y_1 in the graph $\Gamma_{\mathcal{Y}_K}$ (this is possible because $\Gamma_{\mathcal{Y}_K}$ is connected). $y_2, \ldots, y_r$ can be divided into two subgroups: the subgroup $y_2, \ldots, y_{r_1}$ of elements such that each element of $y_{r+1}, \ldots, y_p$ is directly connected to some element of this subgroup and subgroup $y_{1+r_1}, \ldots, y_r$, which is not connected to any element from $y_{1+r}, \ldots, y_p$ as shown in Figure 2. Note that $y_2, \ldots, y_{r_1}$ and $y_{1+r_1}, \ldots, y_r$ may have some connections. Henceforth, we assume that $r > r_1$. It is easy to adapt the proof to the case in which $r = r_1$. Note further that it may be the case that some elements of $y_{r+1}, \ldots, y_p$ are not directly connected to any element of $y_2, \ldots, y_{r_1}$. As B is well connected, they are connected to some elements of $y_{r+1}, \ldots, y_p$, and we can also adapt the proof to this case by adding a layer on our tree, as shown in the right part of Figure 2. We assume henceforth that this is not the case, but again, the proof is similar.

(ii) If S contains y_1, it is easy to prove that $\mathbb{P}(S) \leq \mathcal{L}(S)$ because we simply subtract some ε.

(iii) We have to prove it now for the subset S that does not contain y_1.

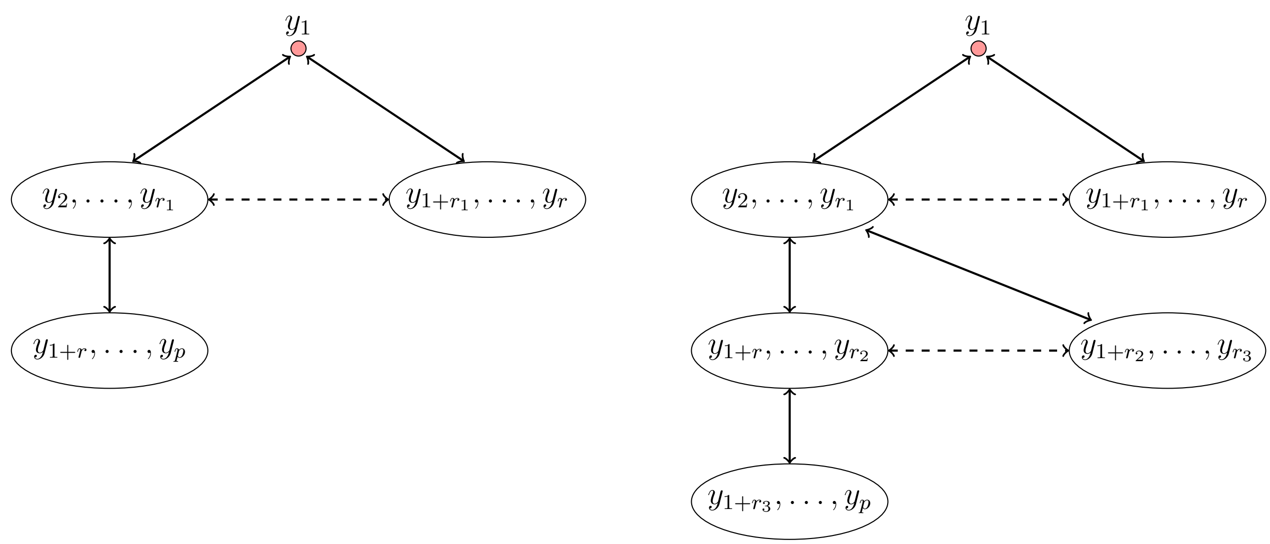


Figure 2: Construction of a tree from elements of B starting from y_1 and an additional layer in the tree if all $y_{1+r}, \ldots, y_p$ may not be connected directly to some element $y_2, \ldots, y_{1+r_1}$.

Let z_2 be an element of $\{y_2, \ldots, y_{r_1}\}$, z_3 be an element of $\{y_{1+r_1}, \ldots, y_r\}$ and z_4 be an element of $\{y_{1+r}, \ldots, y_p\}$. First, y_1 and z_2 are connected. This means that there is at least one region of multiple equilibria that predicts y_1 and z_2 among other outcomes. We call Δ_2 the area of this region. Obviously, we have $\Delta_2 > 0$. Similarly, we have $\mathcal{L}(\{y_1, z_2\}) \leq \mathcal{L}(\{y_1\}) + \mathcal{L}(\{z_2\}) - \Delta_2$ because Δ_2 is counted in both $\mathcal{L}(\{y_1\})$ and $\mathcal{L}(\{z_2\})$ (and there may be other regions of multiple equilibria than the one considered here that predict these outcomes). We do the same for y_1 and z_3 with Δ_3 and the same for z_4 and one element of $\{y_2, \ldots, y_{r_1}\}$ that we call z_2' with Δ_4. z_2' may be z_2 or not. The construction is described in Figure 3.

(a) We have

$$\mathbb{P}(z_2) \leq \mathcal{L}(\{z_2\}) - \Delta_2 + \varepsilon.$$

If $z_2 = y_2$, the inequality expressed above yields the following:

$$\begin{aligned}\mathbb{P}(\{y_2\}) &= \mathcal{L}(\{y_1, y_2\}) - \mathcal{L}(\{y_1\}) + \varepsilon \\ &\leq \mathcal{L}(\{y_1\}) + \mathcal{L}(\{y_2\}) - \Delta_2 - \mathcal{L}(y_1) + \varepsilon.\end{aligned}$$

If $z_2 = y_3$, we can prove it similarly:

$$\mathbb{P}(y_3) = \mathcal{L}(\{y_1, y_2, y_3\}) - \mathcal{L}(\{y_1, y_2\}) + \varepsilon$$

$$\leq \mathcal{L}(\{y_1, y_2\}) + \mathcal{L}(\{y_3\}) - \Delta_2 - \mathcal{L}(\{y_1, y_2\}) + \varepsilon.$$

and so forth (the last inequality holds because there is at least the region of area Δ_2 in multiplicity between $z_2 = y_3$ and $y_1 \in \{y_1, y_2\}$).

(b) Similarly, $\mathbb{P}(z_3) \leq \mathcal{L}(\{z_3\}) - \Delta_3 + \varepsilon$ and $\mathbb{P}(z_4) \leq \mathcal{L}(\{z_4\}) - \Delta_4 + \varepsilon$

(c) Again $\mathbb{P}(z_2, z_3) \leq \mathcal{L}(\{z_2, z_3\}) - \min(\Delta_2, \Delta_3) + 2\varepsilon$, $\mathbb{P}(z_3, z_4) \leq \mathcal{L}(\{z_3, z_4\}) - \min(\Delta_3, \Delta_4) + 2\varepsilon$, $\mathbb{P}(z_2, z_4) \leq \mathcal{L}(\{z_2, z_4\}) - \min(\Delta_2, \Delta_4) + 2\varepsilon$ and $\mathbb{P}(z_2, z_3, z_4) \leq \mathcal{L}(\{z_2, z_3, z_4\}) - \min(\Delta_2, \Delta_3, \Delta_4) + 3\varepsilon$.

(d) Therefore, if $3\varepsilon < \min_{i \in \{2,3,4\}} \Delta_i$, then $\mathbb{P}(S) \leq \mathcal{L}(S)$ for every $S \subset \{z_2, z_3, z_4\}$.

It is straightforward to extend the argument for any subset that contains elements of the type (z_2, z_3, z_4). We need to choose ε such that $(p-1)\varepsilon < \min_{S \in S^{(K)}} \Delta_S^{(K)}(\theta)$.

We therefore have $\mathbb{P}(S) \leq \mathcal{L}(S)$ for every $S \subset B$.

(iv) As $\mathbb{P}(S) \leq \mathcal{L}(S)$ for every $S \subset B$, it is easy to see that this is also satisfies for any union $S \cup C$, where $C \subseteq \mathcal{Y}_K \backslash B$. We still have to prove that the inequalities $\mathbb{P}(S) \leq \mathcal{L}(S)$ are satisfied for $S = B \cup C$, where $C \subseteq \mathcal{Y}_K \backslash B$. We will build a similar tree for $\mathcal{Y}_K \backslash B$. Select $y_{p+1} \in \mathcal{Y}_K \backslash B$. If C contains y_{p+1}, checking the inequality is straightforward. Now, we have to prove this when C does not contain y_{p+1}. The proof is similar to that above.

(v) Find s such that y_{p+1} is directly connected to each $y_{p+2}, \ldots, y_s \in \mathcal{Y}_K \backslash B$ in graph $\Gamma_{\mathcal{Y}_K \backslash B}$ (this is possible because $\mathcal{Y}_K \backslash B$ is connected in $\Gamma_{\mathcal{Y}_K \backslash B}$). $y_{p+2}, \ldots, y_s$ can be divided into two subgroups: the subgroup $y_{p+2}, \ldots, y_{s_1}$ such that each outcome $y_{1+s}, \ldots, y_{d_K}$ is directly connected to some element of this subgroup and the subgroup $y_{1+s_1}, \ldots, y_s$, which is not connected to any element from $y_{1+s}, \ldots, y_{d_K}$. Note further that not all $y_{1+s}, \ldots, y_{|\mathcal{Y}_K|}$ may be connected directly to some element $y_{p+2}, \ldots, y_{s_1}$, but if not, then we will only have an additional layer in the tree, and the proof can easily be modified to any additional layer. Two alternative, simplified trees are also built similar to the construction above (see Figure 4). If $\varepsilon < \min_{S \in S^{(K)}} \Delta_S^{(K)}(\theta)$, a similar argument to that above proves that the inequalities are satisfied for any C.

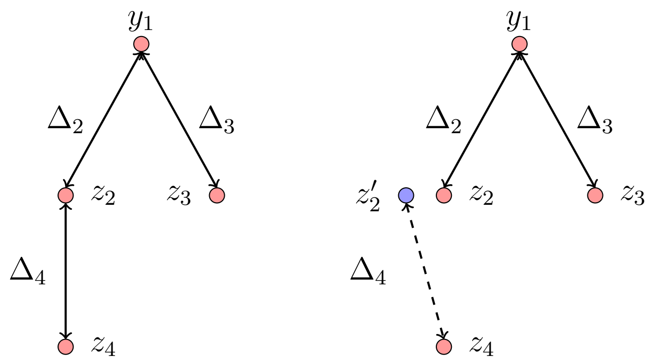


Figure 3: Two simplified trees.

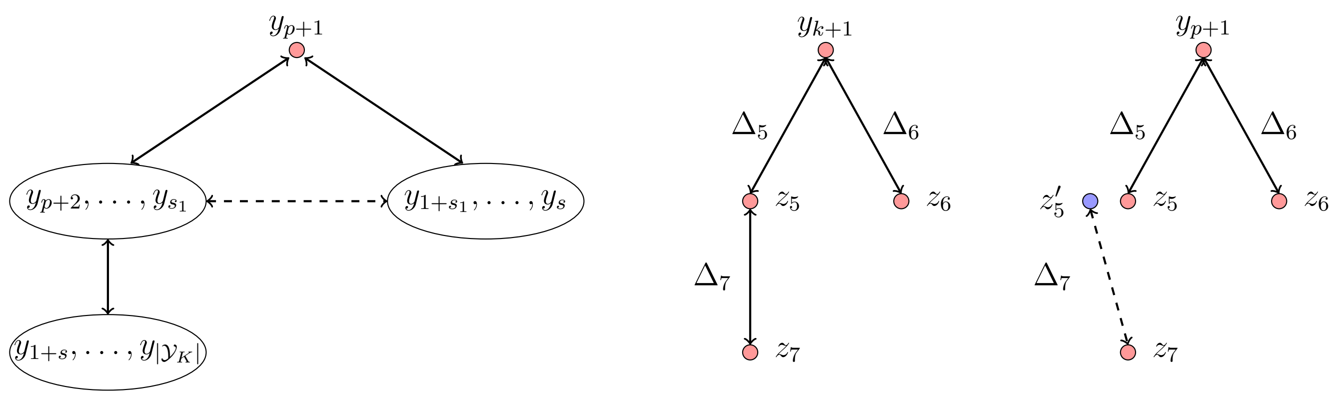


Figure 4: Construction of a tree from elements of $\mathcal{Y}_K \backslash B$ starting from y_{p+1} and two simplified trees.

A.7 Constructing the Core Determining Class

Proposition 7 permits to build an algorithm to construct the core determining class. The algorithm is decomposed in four steps:

1. First, we collect the subsets $B \subseteq \mathcal{Y}_K$ such that B is not connected in Γ_B. We call this collection $\mathcal{D}_K$.

2. Second, we define $\mathcal{D}_K^* = \left\{\mathcal{Y}_K \backslash C : C \in \mathcal{D}_K\right\}$, which is the collection of non well connected subsets of $\mathcal{Y}_K$. As a matter of fact, for any $B \in \mathcal{D}_K^*$, there exists $C \in \mathcal{D}_K$, such that $B = \mathcal{Y}_K \backslash C$ and C, due to the first step, is not connected in Γ_C.

3. Third, we define $\Omega_K = \mathcal{P}^*(\mathcal{Y}_K) \backslash \mathcal{D}_K^*$, which gathers all well connected subsets of $\mathcal{Y}_K$.

4. Finally, we define $\Omega = \left\{\Omega_K : K = 1, \dots, N\right\}$, the well connected subsets of $\mathcal{Y}$.

A.7.1 The algorithm details of the first step.

The first step is the one which needs more details.

Find $\mathcal{D}_K = \left\{B \subseteq \mathcal{Y}_K : \ B \text{ is not connected in } \Gamma_B\right\}$. For simplification, we denote by $\mathcal{P}^*(C)$, for any set C, the collection of all non empty subsets of C and by an abuse of notation $\mathcal{P}^*(\mathcal{C})$, for any collection of sets $\mathcal{C}$, the collection of $\mathcal{P}^*(C)$ for all the elements of $\mathcal{C}$. We also define the concatenation $\oplus$ of two collections C and B is defined as

$$C \oplus B = \bigcup_{c \in C} \left\{c \cup b : b \in B\right\}$$

For example,

$$\left\{\{y_1^{(K)}\}; \{y_2^{(K)}\}\right\} \oplus \left\{\{y_3^{(K)}\}; \{y_4^{(K)}\}\right\} = \left\{\{y_1^{(K)}, y_3^{(K)}\}; \{y_1^{(K)}, y_4^{(K)}\}; \{y_2^{(K)}, y_3^{(K)}\}; \{y_2^{(K)}, y_4^{(K)}\}\right\}.$$

We denote by $S^{(K)}(h)$, the elements of $S^{(K)}$ with h outcomes. The intuition behind the algorithm is to start from the pairs which are not in multiple equilibria and to extend the sequence by sequentially increasing the set with k tuples k increasing from 3 to d_K.

- For any j, consider the set $C_j^{(K)}$ of outcomes which are not in multiple equilibria with $y_j^{(K)}$ and call $C_j^{+(K)}$ the union of $\{y_j^{(K)}\}$ and $C_j^{(K)}$. We now prove that $C_j^{+(k)}$ is the largest subset of $\mathcal{Y}_K$ such that $y_j^{(K)}$ is an isolated node in the graph generated by itself.

  Proof. If $y_j^{(K)}$ would be connected to one another node in the graph generated by $C_j^{+(k)}$, there would exist outcomes $y_{i_1}^{(K)}, \ldots, y_{i_m}^{(K)} \in C_j^{(K)}$ such that $y_j^{(K)}, y_{i_1}^{(K)}, \ldots, y_{i_m}^{(K)}$ are in multiplicity. Following the characterization of the multiple equilibria in Proposition 1, we define the values n_1 of indices where firms always play 1 across all the outcomes, n_0 the indices of the firms which always play 0, and n_s the indices of firms which switch. The series of outcomes gathering all the possible switching values, there exist at least one outcome $y_{i_p}^{(K)}$ among $y_{i_1}^{(K)}, \ldots, y_{i_m}^{(K)} \in C_j^{(K)}$ which differentiate from $y_j^{(K)}$ only from two switcher firms, one switching from 0 to 1 and one switching from 1 to 0 when going from $y_j^{(K)}$ to $y_{i_p}^{(K)}$. This is in contradiction with the definition of $C_k^{(K)}$ which collects all outcomes which can't be in multiplicity with $y_j^{(K)}$.

  There $y_j^{(K)}$ is isolated in $C_j^{+(k)}$ and any other outcome outside this set being in multiple equilibria with $y_j^{(K)}$ can't be added to this set. ■

  Therefore, we initialize our construction of the set $\mathcal{D}_K$ by collecting across j all subsets of $\mathcal{Y}_K$ which contain $y_j^{(K)}$ and any part of $C_j^{(K)}$:

$$\mathcal{S}_{K,1} = \bigcup_{j=1}^{d_K} \left\{ y_j^{(K)} \oplus \mathcal{P}^* \left(C_j^{(K)} \right) \right\}.$$

- Now, we extend the construction. We first consider any pair $\left\{y_i^{(K)}, y_j^{(K)}\right\}$ in multiplicity. We can show that $C = \left\{y_i^{(K)}, y_j^{(K)}\right\} \cup \left\{C_i^{(K)} \cap C_j^{(K)}\right\}$, is not connected in Γ_C. The proof is similar than above. We can therefore augment $\mathcal{S}_{K,1}$ by all the possible combinations of the previous type:

$$\mathcal{S}_{K,2} = \mathcal{S}_{K,1} \cup \left\{ \bigcup_{i,j \text{ s.t. } \{y_i^{(K)}, y_j^{(K)}\} \in S_K(2) \text{ and } C_i^{(K)} \cap C_j^{(K)} \neq \emptyset} \left\{ \{y_i^{(K)}, y_j^{(K)}\} \oplus \mathcal{P}^* \left(C_i^{(K)} \cap C_j^{(K)} \right) \right\} \right\}.$$

- and so on, with triples, ...until $h = d_K$.

$$\mathcal{S}_{K,h} = \mathcal{S}_{K,h-1} \cup \left\{ \bigcup_{\left\{a\in\mathcal{P}^*(S_K(h))\setminus\mathcal{P}^*(S_K(h-1)):\bigcap_{j,y_j^{(K)}\in a} C_j^{(K)}\neq\emptyset\right\}} \left\{ a \oplus \mathcal{P}^* \left(\bigcap_{j,y_j^{(K)}\in a} C_j^{(K)} \right) \right\} \right\}.$$

Now $\mathcal{S}_{K,d_K} = \mathcal{D}_K$. Take any set B not connected in Γ_B. There exists a component C of B in Γ_B. Define n_1, n_0 and n_s like above (see Proposition 1), this set is picked in step $h = \binom{n_s}{K-n_1}$.

A.7.2 The Core Determining Class for $N = 4$

We now apply the previous construction for the entry game with four players. First, for $K \neq 2$, any subset of $\mathcal{Y}_K$ is in the core determining class because all series of outcomes are in multiplicity. Therefore, we only detail the case $K = 2$. There are six outcomes in $B_2(\theta)$.

$$\begin{aligned} y_1^{(2)} &= (1,1,0,0)^\top, \\ y_2^{(2)} &= (1,0,1,0)^\top, \\ y_3^{(2)} &= (1,0,0,1)^\top, \\ y_4^{(2)} &= (0,1,1,0)^\top, \\ y_5^{(2)} &= (0,1,0,1)^\top, \\ y_6^{(2)} &= (0,0,1,1)^\top. \end{aligned}$$

First, we apply proposition 1 to find all the elements of $S^{(2)}$, the outcomes in multiple equilibria

$$\begin{aligned} S^{(2)} = \Big\{ & \{y_1^{(2)}, y_2^{(2)}\}, \{y_1^{(2)}, y_3^{(2)}\}, \{y_1^{(2)}, y_4^{(2)}\}, \{y_1^{(2)}, y_5^{(2)}\}, \{y_2^{(2)}, y_3^{(2)}\}, \{y_2^{(2)}, y_4^{(2)}\}, \{y_2^{(2)}, y_6^{(2)}\}, \{y_3^{(2)}, y_5^{(2)}\}, \\ & \{y_3^{(2)}, y_6^{(2)}\}, \{y_4^{(2)}, y_5^{(2)}\}, \{y_4^{(2)}, y_6^{(2)}\}, \{y_5^{(2)}, y_6^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_4^{(2)}\}, \{y_1^{(2)}, y_3^{(2)}, y_5^{(2)}\}, \\ & \{y_1^{(2)}, y_4^{(2)}, y_5^{(2)}\}, \{y_2^{(2)}, y_3^{(2)}, y_6^{(2)}\}, \{y_2^{(2)}, y_4^{(2)}, y_6^{(2)}\}, \{y_3^{(2)}, y_5^{(2)}, y_6^{(2)}\}, \{y_4^{(2)}, y_5^{(2)}, y_6^{(2)}\}, \\ & \{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}, y_4^{(2)}, y_5^{(2)}, y_6^{(2)}\} \Big\} \end{aligned}$$

- It happens that, for any j, $y_j^{(2)}$ and $y_{7-j}^{(2)}$ are never in multiplicity. So, following our algorithm, we have

$$\begin{aligned} \mathcal{S}_{2,1} &= \bigcup_{j=1}^{6} \left\{ y_j^{(2)} \oplus \mathcal{P}^* \left(y_{7-j}^{(2)} \right) \right\} \\ &= \left\{ \{y_1^{(2)}, y_6^{(2)}\}, \{y_2^{(2)}, y_5^{(2)}\}, \{y_3^{(2)}, y_4^{(2)}\} \right\}. \end{aligned}$$

- For $h=2$, there is no pair $\{y_i^{(2)}, y_j^{(2)}\}$ in $S^{(2)}(2)$ such that $\{C_i^{(2)} \cap C_j^{(2)}\} \neq \emptyset$.

- For $h=3$, this is the same ; there is no 3-tuple $\{y_i^{(2)}, y_j^{(2)}, y_l^{(2)}\}$ in $S^{(2)}(3)$ such that $\{C_i^{(2)} \cap C_j^{(2)} \cap C_l^{(2)}\} \neq \emptyset$.

- For $h=4$ or 5, there is no element in $S^{(2)}(h)$.

- Finally for $h=6$, there is only one element, $\{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}, y_4^{(2)}, y_5^{(2)}, y_6^{(2)}\}$. But the intersection of the $C_j^{(2)}$ for all these elements is empty.

Therefore

$$\mathcal{D}_2 = \Big\{\{y_1^{(2)}, y_6^{(2)}\}, \{y_2^{(2)}, y_5^{(2)}\}, \{y_3^{(2)}, y_4^{(2)}\}\Big\}$$

and

$$\begin{aligned}\mathcal{D}_2^* &= \Big\{\{y_1^{(2)}, y_6^{(2)}\}, \{y_2^{(2)}, y_5^{(2)}\}, \{y_3^{(2)}, y_4^{(2)}\} \backslash D, D \in \mathcal{D}_2\Big\}, \\ &= \Big\{\{y_2^{(2)}, y_3^{(2)}, y_4^{(2)}, y_5^{(2)}\}, \{y_1^{(2)}, y_3^{(2)}, y_4^{(2)}, y_6^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_5^{(2)}, y_6^{(2)}\}\Big\}.\end{aligned}$$

Among all the non empty subparts of $\mathcal{Y}_2$, i.e. 63 sets, only 3 are not in the core determining class. For example, Figure 5 draws the graph $(V_{\{y_3^{(2)}, y_4^{(2)}, y_5^{(2)}, y_6^{(2)}\}}, E_{\{y_3^{(2)}, y_4^{(2)}, y_5^{(2)}, y_6^{(2)}\}})$ from the knowledge of S_2 (there is no link between $y_3^{(2)}$ and $y_4^{(2)}$ because they don't occur in multiplicity involving only outcomes $\{y_3^{(2)}, y_4^{(2)}, y_5^{(2)}, y_6^{(2)}\}$). This graph is clearly connected, so $\{y_1^{(2)}, y_2^{(2)}\}$ is well connected and is part of the core determining class. A contrario, $\{y_3^{(2)}, y_4^{(2)}\}$ is not connected in $\Gamma_{\{y_3^{(2)}, y_4^{(2)}\}}$ because these outcomes are not in multiplicity. Therefore $\{y_1^{(2)}, y_2^{(2)}, y_5^{(2)}, y_6^{(2)}\}$ is not well connected and is not part of the core determining class.

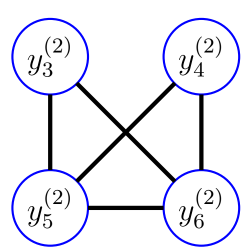


Figure 5: Graph $(V_{\{y_3^{(2)}, y_4^{(2)}, y_5^{(2)}, y_6^{(2)}\}}, E_{\{y_3^{(2)}, y_4^{(2)}, y_5^{(2)}, y_6^{(2)}\}})$ from the multiplicity.

The core determining class is

$$\Big\{\{y_1^{(2)}\},\{y_2^{(2)}\},\{y_3^{(2)}\},\{y_4^{(2)}\},\{y_5^{(2)}\},\{y_6^{(2)}\},\{y_1^{(2)},y_2^{(2)}\},\{y_1^{(2)},y_3^{(2)}\},\{y_1^{(2)},y_4^{(2)}\},\{y_1^{(2)},y_5^{(2)}\},\{y_1^{(2)},y_6^{(2)}\}$$
$$\{y_2^{(2)},y_3^{(2)}\},\{y_2^{(2)},y_4^{(2)}\},\{y_2^{(2)},y_5^{(2)}\},\{y_2^{(2)},y_6^{(2)}\},\{y_3^{(2)},y_4^{(2)}\},\{y_3^{(2)},y_5^{(2)}\},\{y_3^{(2)},y_6^{(2)}\},\{y_4^{(2)},y_5^{(2)}\},\{y_4^{(2)},y_6^{(2)}\}$$
$$\{y_5^{(2)},y_6^{(2)}\},\{y_1^{(2)},y_2^{(2)},y_3^{(2)}\},\{y_1^{(2)},y_2^{(2)},y_4^{(2)}\},\{y_1^{(2)},y_2^{(2)},y_5^{(2)}\},\{y_1^{(2)},y_2^{(2)},y_6^{(2)}\},\{y_1^{(2)},y_3^{(2)},y_4^{(2)}\},\{y_1^{(2)},y_3^{(2)},y_5^{(2)}\}$$
$$\{y_1^{(2)},y_3^{(2)},y_6^{(2)}\},\{y_1^{(2)},y_4^{(2)},y_5^{(2)}\},\{y_1^{(2)},y_4^{(2)},y_6^{(2)}\},\{y_1^{(2)},y_5^{(2)},y_6^{(2)}\},\{y_2^{(2)},y_3^{(2)},y_4^{(2)}\},\{y_2^{(2)},y_3^{(2)},y_5^{(2)}\},\{y_2^{(2)},y_3^{(2)},y_6^{(2)}\}$$
$$\{y_2^{(2)},y_4^{(2)},y_5^{(2)}\},\{y_2^{(2)},y_4^{(2)},y_6^{(2)}\},\{y_2^{(2)},y_5^{(2)},y_6^{(2)}\},\{y_3^{(2)},y_4^{(2)},y_5^{(2)}\},\{y_3^{(2)},y_4^{(2)},y_6^{(2)}\},\{y_3^{(2)},y_5^{(2)},y_6^{(2)}\},\{y_4^{(2)},y_5^{(2)},y_6^{(2)}\}$$
$$\{y_1^{(2)},y_2^{(2)},y_3^{(2)},y_4^{(2)}\},\{y_1^{(2)},y_2^{(2)},y_3^{(2)},y_5^{(2)}\},\{y_1^{(2)},y_2^{(2)},y_3^{(2)},y_6^{(2)}\},\{y_1^{(2)},y_2^{(2)},y_4^{(2)},y_5^{(2)}\},\{y_1^{(2)},y_2^{(2)},y_4^{(2)},y_6^{(2)}\}$$
$$\{y_1^{(2)},y_3^{(2)},y_4^{(2)},y_5^{(2)}\},\{y_1^{(2)},y_3^{(2)},y_5^{(2)},y_6^{(2)}\},\{y_1^{(2)},y_4^{(2)},y_5^{(2)},y_6^{(2)}\},\{y_2^{(2)},y_3^{(2)},y_4^{(2)},y_6^{(2)}\},\{y_2^{(2)},y_3^{(2)},y_5^{(2)},y_6^{(2)}\}$$
$$\{y_2^{(2)},y_4^{(2)},y_5^{(2)},y_6^{(2)}\},\{y_3^{(2)},y_4^{(2)},y_5^{(2)},y_6^{(2)}\},\{y_1^{(2)},y_2^{(2)},y_3^{(2)},y_4^{(2)},y_5^{(2)}\},\{y_1^{(2)},y_2^{(2)},y_3^{(2)},y_4^{(2)},y_6^{(2)}\}$$
$$\{y_1^{(2)},y_2^{(2)},y_3^{(2)},y_5^{(2)},y_6^{(2)}\},\{y_1^{(2)},y_2^{(2)},y_4^{(2)},y_5^{(2)},y_6^{(2)}\},\{y_1^{(2)},y_3^{(2)},y_4^{(2)},y_5^{(2)},y_6^{(2)}\},\{y_2^{(2)},y_3^{(2)},y_4^{(2)},y_5^{(2)},y_6^{(2)}\}$$
$$\{y_1^{(2)},y_2^{(2)},y_3^{(2)},y_4^{(2)},y_5^{(2)},y_6^{(2)}\}\Big\}$$

A.8 The geometric selection procedure

The geometric selection procedure consists in, first determining the local extreme point and, then, deriving the supporting hyperplanes at this extreme point. We adopt the convention of $e_{s_1 i_1,\dots,s_k i_k}$ is the vector where the component i_j is 1 if $s_j = +1$ and -1 if $s_j = -1$.

Determining the local extreme point The procedure to determine the local extreme point is the following:

(1) Pick K and select the subvector $P_0^{(K)}$. This is a vector in a space of dimension d_K.

(2) For each component i, $i = 1, \dots, d_K$, calculate the support function in direction e_i and e_{-i}. Calculate the width in direction e_i, i.e. $D_i = \delta^*(e_i; B_K(\theta)) + \delta^*(e_{-i}; B_K(\theta))$. Calculate the distance to the center point of the cube along the axis of component i: $x_i = \left\{ e_i^\top P_0^{(K)} - \delta^*(e_i; B_K(\theta)) + \frac{D_i}{2} \right\}$.

(3) Pick the coordinate i_1 of the highest values of $|x_i|$. If it is $x_{i_1} > 0$, i_1 is the highest index of the local extreme point, i.e. the local vertex is related to the order $i_1 ? \dots ?$ where the remaining indices need to be found ; otherwise i_1 is the lowest index, i.e. the local vertex is related to the order $? \dots ? i_1$.

(a) Assume $x_{i_1} > 0$. Pick e_{i_1}, and construct the orthogonal projection of $P_0^{(K)}$, $P_{i_1}^{(K)}$, onto the facet F_{i_1}, $x_{i_1} = \delta^*(e_{i_1}; B_K(\theta))$. Then restart Step 2 with now the second index, i'. For each $i' \neq i_1$, take $e_{i_1,i'}$ and $e_{i_1,-i'}$. Compute the width of the intersection of the facet and the set $B_K(\theta)$:

$$D_{i_1,i'} = \delta^*(e_{i_1,i'}; B_K(\theta)) - \delta^*(e_{i_1}; B_K(\theta)) + \delta^*(e_{i_1,-i'}; B_K(\theta)) - \delta^*(e_{i_1}; B_K(\theta)).$$

Calculate the distance to the center of the new cube which contains this intersection:

$$x_{i_1,i'} = \left(e_{i_1,i'}^\top P_{i_1}^{(K)} - \delta^*(e_{i_1,i'}; B_K(\theta)) + \frac{D_{i_1,i'}}{2} \right).$$

Pick the coordinate i_2 of the highest values of $|x_{i_1,i'}|$.

(b) If now $x_{i_1} \leq 0$. Pick e_{i_1}, and construct the orthogonal projection of $P_0^{(K)}$, $P_{i_1}^{(K)}$, onto the facet F_{-i_1}, $x_{i_1} = -\delta^*(e_{-i_1}; B_K(\theta))$. Then restart Step 2 with now the second index, i'. For each $i' \neq i_1$, take $e_{-i_1,i'}$ and $e_{-i_1,-i'}$. Compute the width of the intersection of the facet and the set $B_K(\theta)$:

$$D_{-i_1,i'} = \delta^*(e_{-i_1,i'}; B_K(\theta)) - \delta^*(e_{-i_1}; B_K(\theta)) + \delta^*(e_{-i_1,-i'}; B_K(\theta)) - \delta^*(e_{-i_1}; B_K(\theta)).$$

Calculate the distance to the center of the new cube which contains this intersection:

$$x_{-i_1,i'} = \left(e_{-i_1,i'}^\top P_{i_1}^{(K)} - \delta^*(e_{-i_1,i'}; B_K(\theta)) + \frac{D_{-i_1,i'}}{2} \right).$$

Pick the coordinate i_2 of the highest values of $|x_{-i_1,i'}|$.

(4) Repeat loop 2 and 3 until determining $i_1 i_2 \ldots i_{d_K}$. This is our local extreme point.

(5) Do this procedure for all the values K. Collect the local extreme points accordingly.

Finding the facets at one extreme point Assume the extreme point is $E_{i_1,i_2,\ldots,i_{d_K}}^{(K)}(\theta)$. We now want to determine the facets of $B_K(\theta)$ at this extreme point. The algorithm, detailed below, is based on the idea that, when multiple equilibria does not exist between a series of outcomes, the corresponding indices can be swapped if they are in consecutive ranks without changing the point $E_{i_1,i_2,\ldots,i_{d_K}}^{(K)}(\theta)$ in the space. It increases the number of inequalities that are binding at this extreme point.

1: Start with $\mathcal{L}_K = \big\{\{y_{i_1}^{(K)}\}\big\}$ and set $k = 2$.

k: Find the largest m such that $\big\{y_{i_k}^{(K)}, y_{i_{k+1}}^{(K)}, \ldots, y_{i_m}^{(K)}\big\}$ are not in multiplicity with elements in $\big\{y_{i_j}^{(K)}\big\}$, $j \geq k$. Note $\big\{y_{i_k}^{(K)}, y_{i_{k+1}}^{(K)}, \ldots, y_{i_m}^{(K)}\big\}$ can be in multiplicity with elements in $\big\{y_{i_j}^{(K)}\big\}$, $j \leq k-1$.

$$\mathcal{L}_K = \mathcal{L}_K \cup \Big\{ \big\{y_{i_1}^{(K)}, y_{i_2}^{(K)}, \ldots, y_{i_{k-1}}^{(K)}\big\} \oplus \mathcal{P}^*\big\{y_{i_k}^{(K)}, y_{i_{k+1}}^{(K)}, \ldots, y_{i_m}^{(K)}\big\} \Big\}.$$

Then, update to $k+1$.[21]

R: Repeat the previous step for $k = 2, \ldots, d_K$ steps and find

$$\mathcal{L}_K = \mathcal{L}_K \cap \Omega_K.$$

$\mathcal{L}_K$ can be converted into equivalent support directions. Any element C_K of $\mathcal{L}_K$ yields to the direction e_{C_K} following Equation (10).

We provide a simple illustration of this algorithm in section A.9. The local geometry of set $A(\theta)$ in the extreme point considered is

$$\mathcal{L} = \Big\{ \mathcal{L}_K : K = 0, 1, \ldots, N \Big\}. \tag{A.9}$$

Note that the composition of $\mathcal{L}$ is specific to each extreme point.

A.9 Determining the number of facets for $B_2(\theta)$ when K=4

Following the previous section, we now illustrate how to determine the number of facets in a given extreme point. Consider, for example, the extreme point $E_{1,2,3,4,5,6}^{(2)}(\theta)$ of $B_2(\theta)$. We now determine the number of facets. We know that it is at least 6 but, due to the fact that some outcomes are not in multiplicity, we know that this point is also the same point than $E_{1,2,4,3,5,6}^{(2)}(\theta)$. The procedure determines that. We show that, for this point:

[21] We define the concatenation $\oplus$ of two collections C and B as

$$C \oplus B = \bigcup_{c \in C} \Big\{ c \cup b : b \in B \Big\}$$

and $\mathcal{P}^*$ is defined in Appendix A.7.1.

$$\mathcal{L}_K = \Big\{\{y_1^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_4^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}, y_4^{(2)}\}, \\ \{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}, y_4^{(2)}, y_5^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}, y_4^{(2)}, y_5^{(2)}, y_6^{(2)}\}\Big\}$$

It means that the inequalities which are binding in $E^{(2)}_{1,2,3,4,5,6}(\theta)$ are based on the following directions (that should be completed by zeros accordingly to give direction in $\mathbb{R}^{2^4}$): $e_1 = (1,0,0,0,0,0)^\top$, $e_{1,2} = (1,1,0,0,0,0)^\top$, $e_{1,2,3} = (1,1,1,0,0,0)^\top$, $e_{1,2,4} = (1,1,0,1,0,0)^\top$, $e_{1,2,3,4} = (1,1,1,1,0,0)^\top$, $e_{1,2,3,4,5} = (1,1,1,1,1,0)^\top$ and $e_{1,2,3,4,5,6} = (1,1,1,1,1,1)^\top$.

We now follow the steps of the algorithm introduced in section 4.3.1.

(1) Set $\mathcal{L}_K = \big\{\{y_1^{(2)}\}\big\}$.

(2) At step 2, find the largest m such that $\{y_2^{(2)}, y_3^{(2)}, \ldots, y_m^{(2)}\}$ are not in multiplicity even with outcomes in $\{y_2^{(2)}, y_3^{(2)}, \ldots, y_6^{(2)}\}$. Since $y_2^{(2)}$ and $y_3^{(2)}$ are in multiplicity, $m = 2$.

$$\begin{aligned}\mathcal{L}_K &= \Big\{\{y_1^{(2)}\}\Big\} \cup \Big\{\Big\{\{y_1^{(2)}\} \oplus \mathcal{P}^*(y_2^{(2)})\Big\} \Big\backslash \mathcal{D}_2^*\Big\} \\ &= \Big\{\{y_1^{(2)}\}\Big\} \cup \Big\{\{y_1^{(2)}, y_2^{(2)}\}\Big\} \\ &= \Big\{\{y_1^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}\}\Big\}\end{aligned}$$

(3) At step 3, we look for the largest m such that $\{y_3^{(2)}, \ldots, y_m^{(2)}\}$ are not in multiplicity even with outcomes in $\{y_3^{(2)}, \ldots, y_6^{(2)}\}$. Since $y_3^{(2)}$ and $y_4^{(2)}$ are not in multiplicity, but $y_3^{(2)}$ and $y_5^{(2)}$ are, $m = 4$.

$$\begin{aligned}\mathcal{L}_K &= \Big\{\{y_1^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}\}\Big\} \cup \Big\{\Big\{\{y_1^{(2)}, y_2^{(2)}\} \oplus \mathcal{P}^*\{y_3^{(2)}, y_4^{(2)}\}\Big\} \Big\backslash \mathcal{D}_2^*\Big\} \\ &= \Big\{\{y_1^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_4^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}, y_4^{(2)}\}\Big\}\end{aligned}$$

(4) At step 4, we look for the largest m such that $\{y_4^{(2)}, \ldots, y_m^{(2)}\}$ are not in multiplicity even with outcomes in $\{y_4^{(2)}, \ldots, y_6^{(2)}\}$. Since $y_4^{(2)}$ and $y_6^{(2)}$ are in multiplicity, $m = 5$.

$$\mathcal{L}_K = \Big\{\{y_1^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_4^{(2)}\}, \{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}, y_4^{(2)}\}, \\ \{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}, y_4^{(2)}, y_5^{(2)}\}\Big\}$$

(5) Finally, add $\mathcal{Y}_2 = \left\{\{y_1^{(2)}, y_2^{(2)}, y_3^{(2)}, y_4^{(2)}, y_5^{(2)}, y_6^{(2)}\}\right\}$.

Observe that this does not correspond to the extreme point with the maximum number of facets. This happens in $E^{(2)}_{1,3,4,2,5,6}(\theta)$. In this case, we can start again the algorithm to determine that there are 8 facets defined by the following directions: $e_1 = (1,0,0,0,0,0)^\top$, $e_{1,3} = (1,0,1,0,0,0)^\top$, $e_{1,4} = (1,0,0,1,0,0)^\top$, $e_{1,3,4} = (1,0,1,1,0,0)^\top$, $e_{1,2,3,4} = (1,1,1,1,0,0)^\top$, $e_{1,3,4,5} = (1,0,1,1,5,0)^\top$, $e_{1,2,3,4,5} = (1,1,1,1,1,0)^\top$ and $e_{1,2,3,4,5,6} = (1,1,1,1,1,1)^\top$.

A.10 Proof of Proposition 8

We do the proof for $K = 1$ and it is similar for $K = 2$, which proves the global result.

Fix θ. The goal is to proof that, if a point $P_0^{(1)}$ does not belong to $B_1(\theta)$, a local selection procedure would detect it.

First, observe that any extreme point is linked to an order between the three possible equilibria. Each extreme point $E^{(1)}_{i_1,i_2,i_3}(\theta)$ has supporting hyperplanes with outer normal vectors, $e^{(1)}_{i_1}$, $e^{(1)}_{i_1,i_2}$ and $e^{(1)}_{i_1,i_2,i_3} = (1,1,1)^\top$.

There are three cases:

- If $P_0^{(1)}$ is outside the cube which contains $B_1(\theta)$. It means at least one of the values x_i is outside a bounded interval $[-D_i/2, D_i/2]$, where $D_i =$ is the width in direction e_i. The highest value of $|x_1|$, $|x_2|$, $|x_3|$ selects a face which separates $B_1(\theta)$ and $P_0^{(1)}$. Assume this is $|x_1|$ and that $x_1 > 0$. The first component of $P_0^{(1)}$ is above the largest value of the first component of any point of $B_1(\theta)$. The local extreme point is $E^{(1)}_{1??}(\theta)$. Whatever the next choice, the direction $e_1^{(1)}$ defines a supporting hyperplane of $B_1(\theta)$ at this extreme point which separates $B_1(\theta)$ and $P_0^{(1)}$. Consequently, $T_\infty(e_1^{(1)};\theta) < 0$.

- If $P_0^{(1)}$ is in the cube but not in $B_1(\theta)$. Whatever the choice of the extreme point of $B_1(\theta)$, the third direction $(1,1,1)^\top$ defines a supporting hyperplane which separates $B_1(\theta)$ and $P_0^{(1)}$. Consequently, $T_\infty(e^{(1)}_{1,2,3};\theta) < 0$. And, so forth for the other possibilities.

- If $P_0^{(1)} \in B_1(\theta)$, any choice of local extreme point is valid because, for any direction q, $T_\infty(q, K;\theta) \geq 0$.

When $P_0^{(1)} \in B_1(\theta)$, the procedure does not reject θ. When $P_0^{(1)} \notin B_1(\theta)$, the procedure does. It is therefore a valid and sharp characterization of $B_1(\theta)$.

A.11 Proof of Proposition 9

Under condition UI, following Lemma 3.1 of Romano and Shaikh (2008), we have

$$\sup_{P\in\mathcal{P}} \sup_{S\in\mathcal{L}} \left| \mathbb{P}\left(\sqrt{M}\left(\mu(P) - \hat{P}_M\right) \in S\right) - \Phi_{\Sigma(P)}(S) \right| \underset{M\to\infty}{\longrightarrow} 0,$$

where $\Phi_\Sigma(\cdot)$ is the cumulative distribution function of the centered multivariate normal distribution with variance Σ, $\mu(P) = \mathbb{E}_P(Y)$ and $\Sigma(P) = \text{diag}(\mu(P)) - \mu(P)\mu(P)^\top$ and $\mathcal{L}$ is a collection of convex sets with zero boundary.

Consider the directions q of $\mathcal{G}$ and relabel them $q_1, \ldots, q_m$. Then, define m convex sets in $\mathbb{R}^{2^N}$, D_1, ...,D_m such that,

$$\forall U \in D_i, \quad \frac{q_i^\top U}{\sqrt{q_i^\top \Sigma q_i}} \leq \min_{j\neq i} \frac{q_j^\top U}{\sqrt{q_j^\top \Sigma q_j}}.$$

Now, we can define, for a given $x \in \mathbb{R}$ the sets $S_1, ..., S_m$ ($S_i \subset D_i$) such that

$$\forall U \in S_i, \quad x \leq \frac{q_i^\top U}{\sqrt{q_i^\top \Sigma q_i}} \leq \min_{j\neq i} \frac{q_j^\top U}{\sqrt{q_j^\top \Sigma q_j}}.$$

Now, we have

$$\begin{aligned} \mathbb{P}\left(\inf_{q\in\mathcal{G}}\left(\sqrt{M}\frac{q^\top(\mu(P) - \hat{P}_M)}{\sqrt{q^\top\Sigma q}}\right) \geq x\right) &= \sum_{i=1}^m \mathbb{P}\left(\sqrt{M}(\mu(P) - \hat{P}_M) \in S_i\right) \\ &\underset{M\to\infty}{\longrightarrow} \sum_{i=1}^m \Phi_{\Sigma(P)}\left(S_i\right) = \sum_{i=1}^m \mathbb{P}(Z \in S_i), \end{aligned}$$

uniformly over $P \in \mathcal{P}$, for $Z \sim \mathcal{N}(0, \Sigma(P))$. Moreover,

$$\begin{aligned} \sum_{i=1}^m \mathbb{P}(Z \in S_i) =& \mathbb{P}\left(Z \in \bigcup_{i=1}^m S_i\right) \\ =& \mathbb{P}\left(\inf_{q\in\mathcal{G}} \frac{q^\top Z}{\sqrt{q^\top \Sigma q}} \geq x\right). \end{aligned}$$

So, uniformly over $P \in \mathcal{P}$,

$$\inf_{q \in \mathcal{G}} \left(\sqrt{M} \frac{q^\top(\mu(P) - \hat{P}_M)}{\sqrt{q^\top \Sigma q}} \right) \xrightarrow[M \to \infty]{d} \inf_{q \in \mathcal{G}} \frac{q^\top Z}{\sqrt{q^\top \Sigma q}}. \tag{A.10}$$

Following, Bontemps et al. (2012), proof of Proposition 10, we can now consider the two different cases:

- If P_0, the true choice probability vector belongs to $A(\theta)$, the set of minimizers of $T_M(q;\theta)$ tends to Q_θ, the set of minimizers of $T_\infty(q,\theta)$. This set may not be reduced to a singleton if P_0 is at the intersection of at least two facets. Therefore,

$$\begin{aligned}
\xi_M(\theta) &= \sqrt{M} \min_{q \in \mathcal{G}} \frac{T_M(q;\theta)}{\sqrt{q^\top \Sigma q}} \\
&= \sqrt{M} \inf_{q \in \mathcal{G}} \frac{T_M(q;\theta) - T_\infty(q;\theta) + T_\infty(q;\theta)}{\sqrt{q^\top \Sigma q}} \\
&= \inf_{q \in \mathcal{G}} \frac{\sqrt{M}(q^\top(P_0 - \hat{P}_M)) + \sqrt{M} T_\infty(q;\theta)}{\sqrt{q^\top \Sigma q}} \\
&= \inf_{q \in Q_\theta} \frac{\sqrt{M}(q^\top(P_0 - \hat{P}_M))}{\sqrt{q^\top \Sigma q}}
\end{aligned}$$

  The last equality holds because for any $q \in Q_\theta$, $T_\infty(q;\theta) = 0$ and $q \notin Q_\theta$, $T_\infty(q;\theta) > 0$. So asymptotically, the argmin belongs to Q_θ (remember that $\mathcal{G}$ is discrete). We conclude using the uniform convergence of Equation (A.10).

- If $P_0 \notin A(\theta)$, the value $T_\infty(q,\theta)$ is negative for any direction q. $T_M(q,\theta)$ converges uniformly in q, on the unit sphere, toward a strictly negative value and is therefore bounded away from zero uniformly. The rescaling by $\sqrt{M}$ makes the limit $-\infty$.

Now, we need to consider the fact that Σ is estimated. We need to use the following additional result to replace Σ by $\hat{\Sigma}$ in the proofs above:

$$\sup_{P \in \mathcal{P}} \left\| \hat{\Sigma}(P) - \Sigma(P) \right\| \xrightarrow{P} 0$$

where $\|.\|$ is the component-wise maximum of absolute value of each **element**. This follows from lemma S.7.1 in supplement to Romano and Shaikh (2012).

A.12 Proof of Corollary 11

In the proof of Proposition 9, we show that uniformly over $P \in \mathcal{P}$,

$$\xi_M(\theta) \xrightarrow[M\to\infty]{d} \inf_{(q)\in Q_\theta} \mathcal{N}(0, q^\top \Sigma(P)\, q)$$

if $P_0 \in A(\theta)$. Observe that the distribution depends on θ only through the minimizing set Q_θ, but θ doesn't affect the covariance of the distribution. Define

$$S = \cup_{\theta\in\Theta_I} Q_\theta$$

Since $S \subseteq \mathcal{G}$ the result follows.

A.13 Proof of Proposition 12

Following Proposition 2, we know that any subset of $\mathcal{Y}_1$ of cardinality greater than 2 corresponds to a multiple equilibria region. Consequently, $\Delta_S^{(1)}(\theta)$ for any subset $S \subseteq \mathcal{Y}_1$ is non-zero, and, following Proposition 4, any change in the order gives a different point. Let $i_1, i_2, \ldots, i_{d_1}$ be an order of the coordinates that defines an extreme point $E^{(1)}_{i_1,i_2,\ldots,i_{d_1}}(\theta)$ and $\mathcal{C}^{(1)}_{i_1,\ldots,i_{d_1}}$ be the cone of directions q such that $\delta^*(q; B_1(\theta)) = q^\top E^{(1)}_{i_1,\ldots,i_{d_1}}(\theta)$. Each direction defines a supporting hyperplane (or facet) of $B_1(\theta)$ at $E^{(1)}_{i_1,\ldots,i_{d_1}}(\theta)$.

Any direction q in the cone can be written as

$$q = (q_{i_1} - q_{i_2})e_{i_1} + (q_{i_2} - q_{i_3})e_{i_1,i_2} + \ldots + (q_{i_{d_1-1}} - q_{i_{d_1}})e_{i_1,i_2,\ldots,i_{d_1-1}} + q_{i_{d_1}} e_{i_1,i_2,\ldots,i_{d_1}}.$$

All the coefficients except the last one are positive. The cone is therefore generated by e_{i_1}, e_{i_1,i_2},..., $e_{i_1,i_2,\ldots,i_{d_1-1}}$, $e_{i_1,i_2,\ldots,i_{d_1}}$ or $-e_{i_1,i_2,\ldots,i_{d_1}}$.[22]

In other words, there are only d_1 supporting hyperplanes of $B_1(\theta)$ at this point, and it is sufficient to check the inequalities related to these hyperplanes/facets for a point locally around $E^{(1)}_{i_1,\ldots,i_{d_1}}(\theta)$.

A.14 Proof of Proposition 13

To find the upper bound on the number of facets in any extreme point for $1 < K < N - 1$, we first have to pack the indices which correspond to outcomes which are not in multiplicity.

[22]Remark that $e_{i_1,i_2,\ldots,i_{d_1}} = (1, 1, \ldots, 1)^\top$.

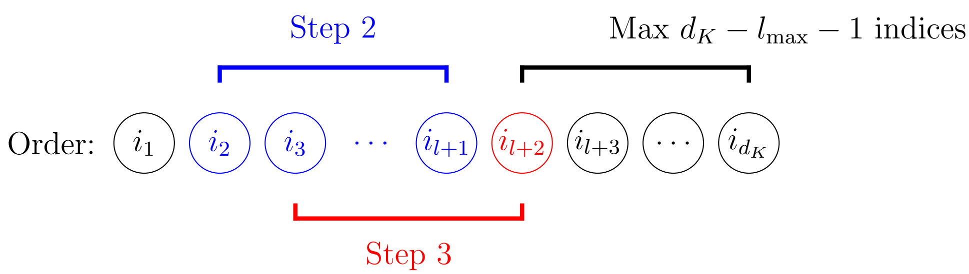
- Let us first define $l_{\max}$ the cardinality of the maximal subset $S \in \mathcal{Y}_K$ such that any pair of S is not in multiplicity. If we collect these indices from the second one, we can switch their order and still have the same point. For example, if $y_{i_2}^{(K)}$ and $y_{i_3}^{(K)}$ are not in multiplicity, the point $E_{i_1,i_2,i_3,???}^{(K)}$ is the same point than $E_{i_1,i_3,i_2,???}^{(K)}$ when the next orders don't change. Consequently, it defines at most one additional outer normal vector following the construction we used earlier in the case $K = 1$. The first outer normal vectors are $e_{i_1}^{(K)}$, $e_{i_1,i_2}^{(K)}$, $e_{i_1,i_3}^{(K)}$, $e_{i_1,i_2,i_3}^{(K)}$, etc. These series of indices are related to $2^{l_{\max}} - 1$ outer normal vector corresponding to all $e_{i_1,i_j}^{(K)}$, $j = 2..l_{\max} + 1$, $e_{i_1,i_j,i_k}^{(K)}$, j, k between 2 and $l_{\max} + 1$ up to $e_{i_1,i_2,\ldots,i_{l_{\max}+1}}^{(K)}$.

- Then we add indices according to the following rule. At each step k, starting from 3, the next index i_m is such that, if possible, $y_{i_k}^{(K)}, \ldots, y_{i_m}^{(K)}$ are not in multiplicity even with all remaining outcomes $\mathcal{Y}_K \backslash \{y_{i_1}^{(K)} \ldots, y_{i_{k-1}}^{(K)}\}$. Otherwise, we pick a random index and we go on. It adds at the maximum (if the added index is not in multiplicity with anybody before) $2^{l_{\max}-1}$ new supporting hyperplanes (as you can switch all orders between k and $k + l_{\max} - 1$, and count only those where the last index is not the last value, i.e. $2^{l_{\max}} - 2^{l_{\max}-1} = 2^{l_{\max}-1}$). See figure 6 below.

- After Step 1, it remains $d_K - l_{\max} - 1$ points after having chosen $i_1, i_2, \ldots, i_{l_{\max}+1}$.

- The maximum number of facets is therefore $\mathcal{L}^*_{\max} = 1 + 2^{l_{\max}} - 1 + (d_K - l_{\max} - 1)(2^{l_{\max}-1}) = 2^{l_{\max}} + (d_K - l_{\max} - 1)(2^{l_{\max}-1})$.

Observe that this bounds is a loose bound and can always be refined by brute force on a case to case basis. Nevertheless, it gives a sufficiently precise estimate of $c(\mathcal{L}^*)$ the true cut off value. When $N = 4$ and $K = 2$, $d_K = 6$ and $l_{max} = 2$ (see Section A.7.2). Applying the formula above we obtain, $\mathcal{L}^*_{\max} = 10$. The true $\mathcal{L}^*$ is equal to 8. However the two cut off values for a level of 5% (see Section 4.3) are $c(8) = -2.51$ and $c(10) = -2.58$. When $N = 5$ and $K = 2$, $d_K = 10$ and $l_{max} = 2$. $\mathcal{L}^*_{\max} = 18$ whereas $\mathcal{L}^* = 15$.



Step 2: we can switch any of these indices. $2^{l}_{\max} - 1$ facets.
Step 3: at max, we can switch any of these indices. $2^{l_{\max}-1}$ additional facets.
Note: l denotes $l_{\max}$ in the circles related to the order.

Figure 6: Counting the maximum number of facets at $E^{(K)}_{i_1,i_2,\ldots,i_{d_K}}(\theta)$.

K	$N=5$	$N=6$	$N=7$	$N=8$	$N=9$
2	2	3	3	4	4
3	2	4	7	8	10
4	1	3	7	14	14
5		1	3	8	14
6			1	4	10
7				1	4
8					1

Table VII: Value of $l_{\max}$ for $N \leq 9$

A.15 The simulation of the critical values

We propose an algorithm to simulate the critical values $c(\mathcal{G},\alpha)$ of Section 4.2. Observe that:

$$\min_{q\in\mathcal{G}} \frac{q^\top Z}{\sqrt{q^\top\hat{\Sigma}q}} = \min_{q\in\mathcal{G}} \frac{(\hat{\Sigma}^{1/2}q)^\top\hat{\Sigma}^{1/2}\tilde{Z}}{\sqrt{(\hat{\Sigma}^{1/2}q)^\top(\hat{\Sigma}^{1/2}q)}} + 0_p(1),$$

where $\tilde{Z}\sim N(0,\hat{\Sigma}^{-1})$. The matrix $\hat{\Sigma}^{1/2}$ rotates the quantities of interest but the direction q which minimizes the quantity are still the same.

1. Draw $\tilde{Z}$ from the normal distribution $\mathcal{N}(0,\hat{\Sigma}^{-1})$. Cut $\tilde{Z}$ in subvectors $\tilde{Z}^{(0)},\tilde{Z}^{(1)},\ldots,\tilde{Z}^{(N)}$.

2. For each $K\in\{0,1,\ldots,N\}$, order $\tilde{Z}_i^{(K)}$, $i=1,\ldots,d_K$ in the increasing order $\tilde{Z}_{i_1}^{(K)},\tilde{Z}_{i_2}^{(K)}, etc.$ The direction q_K which minimizes $q_K^\top\hat{\Sigma}\tilde{Z}^{(K)}/\sqrt{q_K^\top\hat{\Sigma}q_K}$ is among the direction $e_{i_1}^{(K)}$, $e_{i_1,i_2}^{(K)}$, ..., $e_{i_1,i_2,\ldots,i_{d_K}}^{(K)}$. Calculate all the values $q_K^\top\hat{\Sigma}\tilde{Z}^{(k)}/\sqrt{q_K^\top\hat{\Sigma}q_K}$ for all the potential candidates and take the minimum one, called m_K.

3. Take $\underline{m}=\min_{K=0,\ldots,N} m_K$.

4. Repeat the previous steps, $S-1$ times to get S realizations of the distribution of the lower bound and take the α-quantile of this distribution. This is $c(\mathcal{G},\alpha)$.

B The entry game with three players

In this section, we consider our entry game with three firms. The profit of firm i in market m, π_{im} is modeled as:

$$\pi_i = \beta_i + \alpha_i\Big(\sum_{j\neq i} a_j\Big) + \varepsilon_i, \tag{B.11}$$

where a_1 (resp. a_2, a_3) is equal to 1 when $\pi_1 > 0$ (resp. $\pi_2 > 0$, $\pi_3 > 0$), 0 otherwise. The joint distribution of $(\varepsilon_1,\varepsilon_2,\varepsilon_3)$, $F(\cdot;\gamma)$, is assumed to be known up to parameter and θ denote all parameters in the model. We also assume that α's are negative.

B.1 Multiple equilibria regions

There are, in this setup, eight regions of multiple equilibria, which correspond to the set of outcomes expressed in Table VIII. First note that $S^{(1)}=\big\{S_1,S_2,S_3,S_4\big\}$ and $S^{(2)}=\big\{S_5,S_6,S_7,S_8\big\}$.

$N=1$	$S_1 = \begin{pmatrix} 0,0,1 \\ 0,1,0 \end{pmatrix}$	$S_2 = \begin{pmatrix} 0,0,1 \\ 1,0,0 \end{pmatrix}$	$S_3 = \begin{pmatrix} 0,1,0 \\ 1,0,0 \end{pmatrix}$	$S_4 = \begin{pmatrix} 0,0,1 \\ 0,1,0 \\ 1,0,0 \end{pmatrix}$
$N=2$	$S_5 = \begin{pmatrix} 0,1,1 \\ 1,0,1 \end{pmatrix}$	$S_6 = \begin{pmatrix} 0,1,1 \\ 1,1,0 \end{pmatrix}$	$S_7 = \begin{pmatrix} 1,0,1 \\ 1,1,0 \end{pmatrix}$	$S_8 = \begin{pmatrix} 0,1,1 \\ 1,0,1 \\ 1,1,0 \end{pmatrix}$

Table VIII: All multiplicities in pure strategy Nash equilibrium of entry game with 3 players.

B.2 The set of predicted choice probabilities

Recall that the probability of each outcome can be written with the unknown selection mechanism $\eta(\cdot)$. For example,

$$P_{001} = P_1^{(1)}(\theta,\eta) = \Delta_1^{(1)}(\theta) + \sum_{S\in\{S_1,S_2,S_4\}} \int_{\mathcal{R}_S^{(K)}(\theta)} \eta((0,0,1)^\top|\varepsilon,\theta)dF(\varepsilon;\gamma),$$

Let $u_j(S_k)$ be defined like in Section A.4. The set $A(\theta)$ is the collection of points in $\mathbb{R}^8$ that can be written, for some specific choice of weights $u_j(S_k)$:

$$\begin{pmatrix} P_{000} \\ \hline P_{001} \\ P_{010} \\ P_{100} \\ \hline P_{011} \\ P_{101} \\ P_{110} \\ \hline P_{111} \end{pmatrix} = \begin{pmatrix} \Delta_1^{(0)}(\theta) \\ \hline \Delta_1^{(1)}(\theta) + u_1(S_1)\Delta_{S_1}^{(1)}(\theta) + u_1(S_2)\Delta_{S_2}^{(1)}(\theta) + u_1(S_4)\Delta_{S_4}^{(1)}(\theta) \\ \Delta_2^{(1)}(\theta) + u_2(S_1)\Delta_{S_1}^{(1)}(\theta) + u_2(S_3)\Delta_{S_3}^{(1)}(\theta) + u_2(S_4)\Delta_{S_4}^{(1)}(\theta) \\ \Delta_3^{(1)}(\theta) + u_3(S_2)\Delta_{S_2}^{(1)}(\theta) + u_3(S_3)\Delta_{S_3}^{(1)}(\theta) + u_3(S_4)\Delta_{S_4}^{(1)}(\theta) \\ \hline \Delta_1^{(2)}(\theta) + u_1(S_5)\Delta_{S_5}^{(2)}(\theta) + u_1(S_6)\Delta_{S_6}^{(2)}(\theta) + u_1(S_8)\Delta_{S_8}^{(2)}(\theta) \\ \Delta_2^{(2)}(\theta) + u_2(S_5)\Delta_{S_5}^{(2)}(\theta) + u_2(S_7)\Delta_{S_7}^{(2)}(\theta) + u_2(S_8)\Delta_{S_8}^{(2)}(\theta) \\ \Delta_3^{(2)}(\theta) + u_3(S_6)\Delta_{S_6}^{(2)}(\theta) + u_3(S_7)\Delta_{S_7}^{(2)}(\theta) + u_3(S_8)\Delta_{S_8}^{(2)}(\theta) \\ \hline \Delta_1^{(3)}(\theta) \end{pmatrix}, \tag{B.12}$$

with the constraint $\sum_{j|y_j^{(K)}\in S} u_j(S) = 1,\ \ 0 \le u_j(S) \le 1$ for $S \in S^{(K)}$. The partition here refers to different K (number of active firm in any outcome). This partition is very useful as the convex set decomposes into cartesian product of smaller dimension convex set. This convex set only need 18 directions to characterize it.

Figure 7 displays the set $B_1(\theta)$, its outer cube and the inequalities (in red) which are tested in our geometric selection procedure.

B.3 The directions used in the Monte Carlo experiment

Following Proposition 4, we have a closed-form expression for the support function. It is equal to:

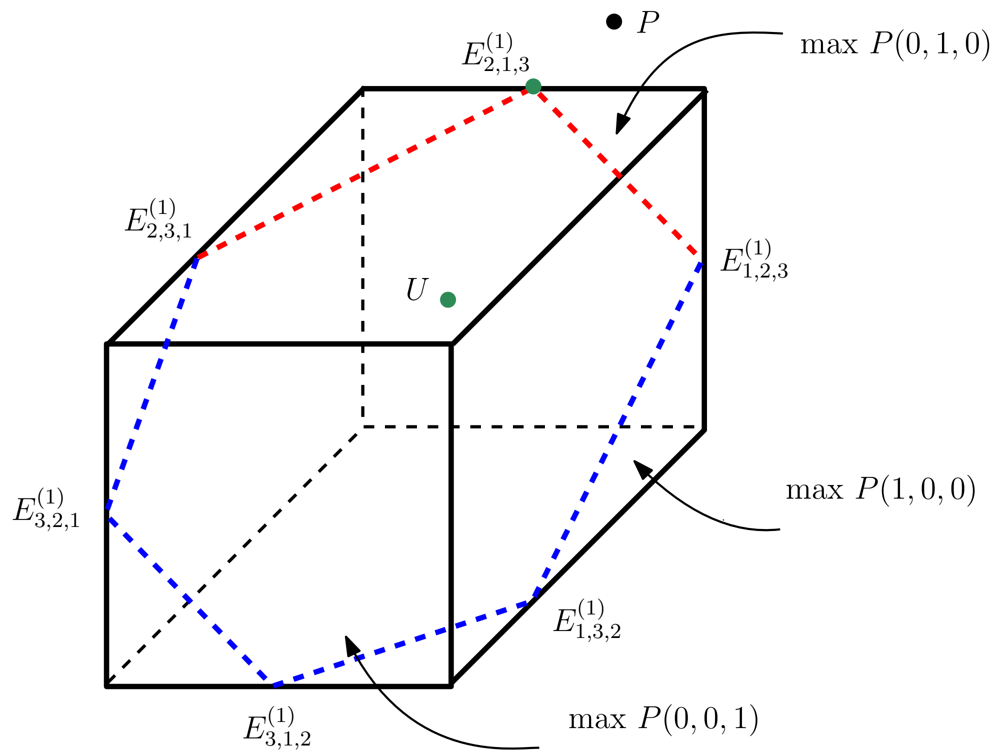


Figure 7: The geometric selection procedure for $B_1(\theta)$.

$$
\begin{aligned}
\delta^*\left(q; A(\theta)\right) = {} & q^\top \Delta(\theta) \\
& + \max(q_2, q_3)\Delta^{(1)}_{S_1}(\theta) + \max(q_2, q_4)\Delta^{(1)}_{S_2}(\theta) + \max(q_3, q_4)\Delta^{(1)}_{S_3}(\theta) + \max(q_2, q_3, q_4)\Delta^{(1)}_{S_4}(\theta) \\
& + \max(q_5, q_6)\Delta^{(2)}_{S_5}(\theta) + \max(q_5, q_7)\Delta^{(2)}_{S_6}(\theta) + \max(q_6, q_7)\Delta^{(2)}_{S_7}(\theta) + \max(q_5, q_6, q_7)\Delta^{(2)}_{S_8}(\theta),
\end{aligned}
$$

where $q = (q_1, \dots, q_8)^\top$ and

$$
\Delta(\theta) = \left(\Delta^{(0)}_1(\theta), \Delta^{(1)}_1(\theta), \Delta^{(1)}_2(\theta), \Delta^{(1)}_3(\theta), \Delta^{(2)}_1(\theta), \Delta^{(2)}_2(\theta), \Delta^{(2)}_3(\theta), \Delta^{(3)}_1(\theta)\right)^\top.
$$

The identified set can be estimated by testing that the point P_0 belongs to $A(\theta)$ or, equivalently, by testing that

$$
\min_{q \in \mathcal{G}} \delta^*\left(q; A(\theta)\right) - q^\top P_0 \geq 0.
$$

The following sets of directions considered for $\mathcal{G}$ are given below. For "Ineq$_1$", we consider 16 inequalities derived from these 16 directions. Each direction is a column of the following set:

$$
\mathcal{G}_1 = \left\{
\begin{array}{cccccccccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}
\right\}.
$$

Similarly, for "Ineq$_2$" we consider also 16 directions:

$$
\mathcal{G}_2 = \left\{
\begin{array}{cccccccccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}
\right\}.
$$

"Ineq$_3$" takes the whole set of inequalities and equalities which define $B_0(\theta)$ $B_1(\theta)$ and $B_3(\theta)$

(the equality related to $B_2(\theta)$ being redundant is dropped):

$$\mathcal{G}_3 = \left\{ \begin{array}{cccccccccccccccccc} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{array} \right\}.$$

"Ineq$_4$" replaces the directions $e_{ij}^{(K)}$ in "Ineq$_3$ by the directions $-e_l^{(K)}$

$$\mathcal{G}_4 = \left\{ \begin{array}{cccccccccccccccccc} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{array} \right\}.$$

C Additional Monte Carlo simulations

We compare our procedure with the results of the GMS procedure proposed by Andrews and Soares (2010) and its refinement by Romano et al. (2012). We use the DGP of the example with three player considered in table IV, i.e, $\beta = 0.35$ and $\alpha_1 = \alpha_2 = \alpha_3 = -0.4$. Table IX displays the mean rejection rate across simulations for a sequence of points on a curve displayed in Figure 8. The sequence of points contains points from the identified set and outside this identified set. First, for $M = 1000$, all rejection rates are fine with values close to 5% at the boundary. Second, the power of our local procedure is better than the GMS procedures because the critical value is smaller.

M=1000				M=2000			
G	L	AS	RSW	G	L	AS	RSW
0.889	0.923	0.914	0.899	0.998	0.999	0.999	0.999
0.826	0.881	0.865	0.841	0.995	0.997	0.997	0.996
0.756	0.820	0.802	0.773	0.984	0.992	0.990	0.988
0.679	0.750	0.731	0.698	0.964	0.978	0.975	0.971
0.595	0.675	0.656	0.616	0.927	0.952	0.950	0.941
0.512	0.594	0.571	0.532	0.875	0.913	0.907	0.897
0.434	0.516	0.492	0.453	0.797	0.854	0.844	0.826
0.365	0.442	0.419	0.383	0.718	0.775	0.764	0.742
0.304	0.375	0.353	0.315	0.625	0.694	0.681	0.657
0.246	0.316	0.294	0.257	0.522	0.603	0.589	0.559
0.196	0.259	0.240	0.206	0.413	0.494	0.479	0.446
0.155	0.209	0.193	0.163	0.319	0.392	0.378	0.351
0.121	0.170	0.157	0.128	0.241	0.300	0.289	0.265
0.095	0.135	0.125	0.101	0.180	0.229	0.218	0.200
0.077	0.112	0.102	0.082	0.127	0.172	0.161	0.145
0.066	0.095	0.086	0.069	0.095	0.124	0.118	0.107
0.056	0.083	0.076	0.057	0.070	0.094	0.088	0.078
0.048	0.073	0.065	0.049	0.052	0.071	0.067	0.060
0.045	0.065	0.059	0.045	0.042	0.058	0.054	0.048
0.043	0.062	0.054	0.043	0.035	0.048	0.045	0.040
0.040	0.059	0.053	0.040	0.033	0.043	0.040	0.036
0.038	0.057	0.049	0.038	0.031	0.041	0.038	0.035
0.037	0.054	0.048	0.037	0.030	0.039	0.036	0.033
0.036	0.054	0.046	0.036	0.030	0.038	0.035	0.031
0.036	0.053	0.045	0.035	0.030	0.038	0.035	0.031
0.036	0.052	0.044	0.035	0.029	0.038	0.035	0.030
0.035	0.052	0.044	0.035	0.029	0.038	0.035	0.030
0.035	0.052	0.043	0.035	0.029	0.038	0.035	0.029
0.035	0.052	0.043	0.035	0.029	0.038	0.035	0.029
0.035	0.052	0.043	0.035	0.029	0.038	0.035	0.028
0.035	0.052	0.044	0.035	0.029	0.038	0.035	0.029
0.035	0.052	0.045	0.035	0.029	0.038	0.035	0.030
0.035	0.052	0.046	0.035	0.029	0.038	0.035	0.030
0.036	0.054	0.048	0.036	0.030	0.038	0.036	0.032
0.043	0.064	0.058	0.044	0.039	0.052	0.048	0.043
0.050	0.075	0.069	0.051	0.056	0.079	0.071	0.064
0.069	0.103	0.094	0.072	0.103	0.140	0.132	0.118
0.152	0.208	0.189	0.158	0.316	0.390	0.375	0.346
0.104	0.145	0.133	0.108	0.196	0.244	0.236	0.215
0.223	0.289	0.269	0.235	0.468	0.541	0.527	0.504
0.300	0.382	0.355	0.314	0.600	0.680	0.669	0.641
0.394	0.471	0.449	0.410	0.729	0.793	0.781	0.758
0.491	0.569	0.548	0.512	0.836	0.882	0.874	0.856
0.600	0.676	0.656	0.618	0.921	0.947	0.945	0.934
0.708	0.774	0.754	0.722	0.963	0.976	0.974	0.970
0.800	0.855	0.841	0.812	0.988	0.993	0.992	0.990
0.877	0.916	0.905	0.888	0.996	0.997	0.997	0.996
0.933	0.958	0.952	0.940	0.998	1.000	0.999	0.999
0.968	0.979	0.976	0.972	1.000	1.000	1.000	1.000
0.985	0.993	0.991	0.988	1.000	1.000	1.000	1.000
0.996	0.998	0.997	0.997	1.000	1.000	1.000	1.000
0.999	1.000	0.999	0.999	1.000	1.000	1.000	1.000

The rejection rates for the points in the identified set are colored in grey.

Table IX: Rejection frequencies for points tested according to different inference methods. See Figure 8 for a plot of the points in the space (α_1, α_2).

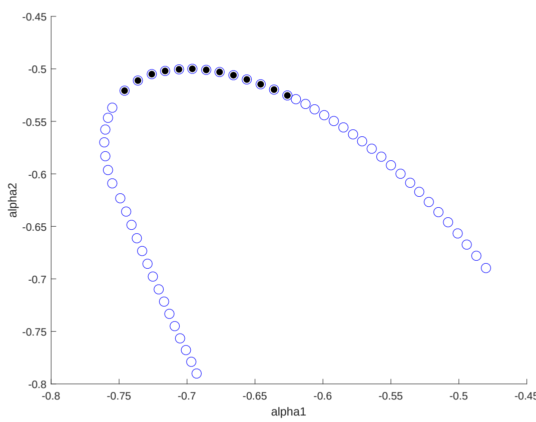


Figure 8: Sequence of points tested (the points of the identified set are colored).