Readings on Moment Inequality Methods

Manu Navjeevan

July 26, 2020

Contents

1 A Practical Method for Testing Many Moment Inequalities; *Yuehao Bai, Andres Santos, Azeem M. Shaikh*

1.1 Introduction

Setup: $\{X_i\}_{i=1}^n$ i.i.d with distribution $P \in P_n$ on R^n. Consider the problem of testing

$$H_0 : P \in \mathbf{P}_{0,n} \quad \text{versus} \quad H_1 : P \in \mathbf{P}_{1,n} \tag{1}$$

where

$$\mathbf{P}_{0,n} \equiv \{P \in \mathbf{P}_n : E_P[X_i] \leq 0\} \tag{2}$$

and $\mathbf{P}_{1,n} = \mathbf{P}_n / \mathbf{P}_{0,n}$. The inequality in 2 is interpreted component wise and $\mathbf{P}_n$ is a large class of possible distributions for the observed data. Indexing both the number of moments p_n and the class of possible distributions by the sample size allows for the number of moments to grow (rapidly) with the sample size n. Goal is to construct test that are uniformply consistent in level; i.e

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_{0,n}} E_P[\phi_n] \leq \alpha \tag{3}$$

A test can be viewed as a function of the data $\phi_n = \phi_n : \mathcal{X}^n \to \{0, 1\}$ where $\mathcal{X}^n$ is generally some subset of $\mathbb{R}^n$ where the data takes its values.

There are a large class of problems in economics in which the number of moments is large. For example, in the entry models as in Cilberto and Tamer (2009) the number of moment inequalities to check is $p_n = o(2^{m+1})$ where m is the number of firms. Apart from Chernozukhov et. al (2019), this has typically been done by limiting $\mathbf{P}_n$ so that the number of moments p_n are small. Canay and Shaikh (2017) provide a detailed review of these tests. This paper focuses on the two step testing procedure of Romano et. al (2014). Test is shown to satisy (3) under assumptions on $\mathbf{P}_n$ that restrict p_n to not depend on n. However, the test is "practical" in that it is computationally feasible even if the number of moments is large. **Paper shows that the test of Romano et. al (2014) continues to satisy (3) for a large class of distributions that permits the number of moments p_n to grow exponentially with the sample size n.**

Theoretical analysis relies on Chernozhukov et. al (2013, 2017) on the high dimensional CLT. This is seminal work. Allen (2018) argues that the test proposed Romano et al. (2014) is more powerful in finite samples then the test proposed by Chernozhukov et al. (2019).

1.2 Main Result

Begin this section by describing the testing procedure in Romano et al. (2014). To do so, best to introduce some further notation. For $1 \leq j \leq p_n$ let $X_{i,j}$ denote the jth component of X_i and set

$$\bar{X}_{j,n} \equiv \frac{1}{n}\sum_{i=1} nX_{i,j} \tag{4}$$

$$S^2_{j,n} \equiv \frac{1}{n}\sum_{i=1}^{n}(X_{i,j} - \bar{X}_{j,n})^2 \tag{5}$$

Can also use the notation $\mu_j(p) \equiv E_P[X_{i,j}]$ and $\sigma^2_j(P) \equiv \text{Var}_P[X_i, j]$ so that (4) and (5) can be expressed as $\mu_j(\hat{P}_n)$ and $\sigma^2_j(\hat{P}_n)$, respectively, where $\hat{P}_n$ is the empirical distribution of $\{X_i\}_{i=1}^n$. Focus on a test that rejects for large values of

$$T_n \equiv \max\left\{\max_{1\leq j\leq p_n} \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}}, 0\right\}$$

In defining critical value, useful to introduce an i.i.d sequence of random variables with distribution $\hat{P}_n$ conditional on $\{X_i\}_{i=1}^n$, which we will denote $X^*_i, i = 1, \ldots, n$. Further define $\bar{X}^*_{j,n}$ and $(S^*_{j,n})^2$ analagously

to before, but substituting in X_i^*. Critical value for T_n is given by

$$\hat{c}_n^{(2)}(1-\alpha+\beta) \equiv \inf \mathcal{S}_n(1-\alpha+\beta) \tag{6}$$

where

$$\mathcal{S}_n(a) \equiv \left\{ c \in \mathbb{R} : \mathbb{P}\left[\max\left\{ \max_j \frac{\sqrt{n}(\bar{X}_{j,n}^* - \bar{X}_{j,n} + \hat{\mu}_{j,n})}{S_{j,n}^*}, 0 \right\} \leq c \;\middle|\; \{X_i\}_{i=1}^n \right] \geq a \right\}$$

Here $\alpha \in (0, 0.5)$ is the nominal level of the test and $\beta \in (0, \alpha)$ and

$$\hat{\mu}_{j,n} \equiv \min\left\{ \bar{X}_{j,n} + \frac{S_{j,n}}{\sqrt{n}} \hat{c}_n^{(1)}(1-\beta), 0 \right\} \tag{7}$$

with

$$\hat{c}_n^{(1)} \equiv \inf\left\{ c \in \mathbb{R} : \mathbb{P}\left[\max_{1 \leq j \leq p_n} \frac{\sqrt{n}(\bar{X}_{j,n} - \bar{X}_{j,n}^*)}{S_{j,n}^*} \leq c \;\middle|\; \{X_i\}_{i=1}^n \right] \geq 1-\beta \right\}$$

The test is then

$$\phi_n^{\text{RSW}} \equiv \mathbb{1}\left\{ T_n \geq \hat{c}_n^{(2)}(1-\alpha+\beta) \right\} \tag{8}$$

Motivating this choice of critical value it is useful to note that the test statistic T_n satisfies

$$T_n = \max\left\{ \max_j \left(\frac{\sqrt{n}(\bar{X}_{j,n} - \mu_j(P))}{S_{j,n}} + \frac{\sqrt{n}\mu_j(P)}{S_{j,n}} \right), 0 \right\} \tag{9}$$

Decomposition highlights that the main impediment in approximating the distribution of T_n is the presence of nuisance parameters $\sqrt{n}\mu_j(P)$ for $1 \leq j \leq p_n$.[1] Though these nuisance parameters cannot be consistently estimated, Romano et al (2014) observe that it may still be possible to construct a suitably valid confidence region for them.

Lemma in Appendix employs Romano insight and high dimensional CLT of Chernozhukov et al. (2017) to show that, under conditions that permit p_n to grow rapidly with the sample size n, $\sqrt{n}\mu_j(P) \leq \sqrt{n}\hat{\mu}_{j,n}$ for all $j \leq p_n$ with pr. approximately no less than $1-\beta$ whenever the null hypothesis in (1) is true. Since T_n is monotonically increasing in the nuisance parameters $\sqrt{n}\mu_j(P)$ for all $1 \leq j \leq p_n$ it follows that, viewed as a funtion of these nuisance parameters, any quantile of T_n is maximized over said confidence region by setting $\sqrt{n}\mu_j(P) = \sqrt{n}\hat{\mu}_{j,n}$ for all j. Then, the critical value $\hat{c}_n^{(2)}(1-\alpha+\beta)$ is a bootstrap estimate of the $1-\alpha+\beta$ quantile of T_n under the "least favorable" nuisance parameter value $\sqrt{n}\mu_j(P) = \sqrt{n}\hat{\mu}_{j,n}$ for all j. The $1-\alpha-\beta$ quantile is employed instead of β to account for that, with pr. appx no greater than β, $\sqrt{n}\mu_j(P) > \sqrt{n}\hat{\mu}_{j,n}$. Analysis of test (8) hinges on following assumption:

Assumption 1. Assume (i) $\{X_i\}_{i=1}^n$ is an i.i.d sample with $X_i \in \mathbb{R}^{p_n}$ and $X_i \sim P \in \mathbf{P}_n$; (ii) $\sigma_j(P) > 0$ for all $1 \leq j \leq p_n$ and $P \in \mathbb{P}_n$; (iii) For $k = 1, 2$, there is a $M_{k,n} < \infty$ such that $E_P[|X_{i,j} - \mu_j(P)|^{2+k}] \leq \sigma_j^{2+k}(P) M_{k,n}^k$ for all $1 \leq j \leq p_n$ and $P \in \mathbf{P}_n$; (iv) There exists a $B_n < \infty$ such that $E_P\left[\max_{1 \leq j \leq p_n} \left|X_{i,j} - \mu_j(P)\right|^4\right] \leq B_n^4$ for all $P \in \mathbf{P}_n$; (v) $(M_{1,n}^2 \vee M_{2,n}^2 \vee B_n^2)\log^{3.5}(p_n n) = o(n^{(1-\delta)/2})$ for some $\delta \in (0,1)$

1(i) formalizes that $\{X_i\}_{i=1}^n$ be an i.i.d sample, while Assumption 1(ii) requires the variance of $X_{i,j}$ to be positive for all $P \in \mathbb{P}_n$ and $1 \leq j \leq p_n$.1(iii) imposes a uniform in P and j bound on the standardized moments of $X_{i,j}$. Condition is a strengthening of the uniform integrability requirements of Romano et al (2014) required so study a setting in which p_n diverges to infinity. Part (iv) bounds the 4th moments of the maximim of $X_{i,j}$. Finally, (v) states the main conditiion governing how fast p_n can grow with n. Under suitable moment restrictions on $X_{i,j}$, p_n may grow exponentially with n. Now ready for main result

[1] I'm not entirely sure why they cannot be consistently estimated. I think this is because we are only partially identified.

Theorem 1. *If Assumption 1 holds,* $\alpha \in (0, \frac{1}{2})$ *and* $0 < \beta < \alpha$*, then* ϕ_n^{RSW} *as defined in (8) satisfies uniform consistency in level as defined in (3)*

The rest of this paper goes through some simulations. It is also just a working paper at the moment. Probably it is best to go through the main proof; but I will print it out and make some notes on this.

2 Inference on Causal and Structural Parameters Using Many Moment Inequalities; *Victor Chernozhukov, Denis Chetverikov, and Kengo Kato (ReStud, 2018)*

Chernozhukov et al. (2018) covers inference on many moment inequalities. Builds on prior work on high dimensional CLT.

2.1 Introduction

In recent years, moment inequalities framework has developed into a powerful tool for inference on causal and structural parameters in partially identified models. Many papers study models with a finite and fixed number of conditional and unconditional moment inequalities. IN practice the number of moment inequalities implied by the model is often large.

Examples of testing (very) many moment inequalities

- Consumer is selecting a bundle of products for purchase and moment inequalities come from revealed preference argument (Pakes, 2010)
- Market structure model of Ciliberto and Tamer (2009), number of moment inequalities equals the number of possible combinations of firms that could potentially enter the market (grows exponentially in the number of firms)
- Dynamic model of imperfect competition of Bajari, Benkard, Levin (2007)m where deviations from optimal policy serve to define many moment inequalities
- Beresteanu, Molchanov, Molinari (2011), Galichon and Henry (2011)[1], Chesher, Rosen, Smolinksi (2013), and Chester and Rosen (2013)

Many examples have inportant in that the many inequalities under consideration are "unstructured", they cannot be viewed as unconditional moment inequalities generated froma small number of conditional inequalities with a low-dimensional conditioning variable. So existing inference methods for conditional moment inequalities, though fruitful in many cases

Formally describing the problem, let $\{X_i\}_{i=1}^n$ be a sequence of i.i.d random vectors in $\mathbb{R}^p$, where $X_i = (X_{i1}, \ldots, X_{ip})^T$, with a common distribution denoted by $\mathcal{L}_X$. For $j \leq p$, we write $\mu_j := \mathbb{E}[X_{1j}]$. Interested in testing the null hypothesis

$$H_0 : \mu_j \leq 0 \text{ for all } j = 1, \ldots, p \tag{1}$$

Against the alternative

$$H_1 : \mu_j > 0 \text{ for some } j = 1, \ldots, p \tag{2}$$

Refer to (1) as the moment inequalities and say the jth moment is satisfied (viloated) if $\mu_j \leq 0$ ($\mu_j > 0$). Paper will allow number of moment inequalities $p \gg n$. Consider a test statistic given by the maximum over p Studentized (t-type) inequality specific statistic. Consider critical values based upon (i) the union bound combined with a moderate deviation inequality for self-normalized sums and (ii) bootstrap methods. Among bootstrap methods, consider multiplier and emppirical bootstrap methods. These are simulation based and computationally more difficult, but take into account correlation structure and yeild lower critical values. SN method is particularly useful for grid search when the researcher is interested in constricting a confidence interval for identified set.

Also consider two-step methods incorporating inequality selection procedutes. Two-step methods get rid of most uninformative inequalities, that is inequalities with $\mu_j < 0$ if μ_j is not too close to 0. Also develop novel three-step methods by incorporating double inequality selection procedures. These are suitable in parametric modesl defined via momoent inequalities and allow to drop weakly informative inequalities in addition to uninformative inequalities.[2]. Results can be used for construction of confidence regions for identifiable parameters in partially identified models defined by moment inequalities. Show that results are asymptotically honest (don't quite know what this means).

[1] This seems like a good place to start reading

[2] Can be extended to nonparametric models as well

Literature testing unconditional moment inequalities is large. See White (2000), Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Rosen (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), Canay (2010), Bugni (2011), Andrews and Jia-Barwick (2012), and Romano, Shaikh, and Wolf (2014).

In this paper we implicitly assume that $X_1, \ldots, X_n$ and p are indexed by n. Mainly interested in the case that $p = p_n \to \infty$ as $n \to \infty$

2.2 Motivating Examples

Section provides examples that motivate the framework where the number of moment inequalities p is large and potentially much larger than the sample size n. In these examples, one actually has many conditional rather than unconditional inequalities. Results cover conditioning as well.

2.2.1 Market Structure Model

Let m denote the number of firms that could potentially enter the market. Let m-tuple $D = (D_1, \ldots, D_m)$ denote entry decisions of these firms. That is, $D_j = 1$ if the firm j enters the market and $D_j = 0$ otherwise. Let $\mathcal{D}$ denote the possible values of D. We have that $|\mathcal{D}| = 2^m$.

Let X and ϵ denote the (exogeneous) characteristics of the market as well as characteristics of the firms that are observed and not observed by the researcher, respectively. The profit of the firm j is given by

$$\pi_j(D, X, \epsilon, \theta)$$

where π_j is known up to a parameter θ. Both X and ϵ are observed by the firms and a Nash Equilibrium is played so that, for each j,

$$\pi_j((D_j, D_{-j}), X, \epsilon, \theta) \geq \pi_j((1 - D_j, D_{-j}), X, \epsilon, \theta)$$

D_{-j} denotes the decisions of all firms excluding the firm j. Then one can find set-valued functions $R_1(d, X, \theta)$ and $R_2(d, X, \theta)$ such that d is the unique equilibrium whenever $\epsilon \in D_1(d, X, \theta)$ and d is *an* equilibrium whenever $\epsilon \in R_2(d, X, \theta)$. In the second case, the probability that the researcher sees d as an equilibrium depends on the equilibrium selection mechanism. Without further information, anything can be in $[0, 1]$. Therefore we have the following bounds

$$\begin{aligned}\mathbb{E}[\mathbb{1}\{\epsilon \in R_1(d, X, \theta\}|X] &\leq \mathbb{E}[\mathbb{1}\{D = d\}|X] \\ &\leq \mathbb{E}[\mathbb{1}\{\epsilon \in R_1(d, X, \theta) \cup R_2(d, X, \theta)\}|X]\end{aligned}$$

Further assuming that the conditional distribution of ϵ given X is known (or known up to a parameter that is part of θ), both the LHS and RHS of these inequalites can be calculated. Denote them $P_1(d, X, \theta)$ and $P_2(d, X, \theta)$, respectively to obtain

$$P_1(d, X, \theta) \leq \mathbb{E}[\mathbb{1}\{D = d\}|X] \leq P_2(d, X, \theta) \tag{3}$$

for all $d \in \mathbb{D}$. These can be used for inference on the parameter θ. Note that the number of inequalities in (3) is $2|\mathcal{D}| = 2^{m+1}$. This is a large number, even if m is moderately large. Moreover, these inequalities are conditional on X. So, they can be transformed into a large and increasing number of unconditional moment inequalities as described above. Also, if the firms have more than two decisions, the number of inequalities will be even larger.

Some other examples are given, but I won't cover them in notes.

2.3 Test Statistic

Begin preparing some notation. Assume that

$$\mathbb{E}[X_{1,j}^2] < \infty, \sigma_j^2 := \mathrm{Var}(X_{1,j}) > 0, j = 1, \ldots, p \tag{4}$$

For $j = 1, \ldots, p$ let $\hat{\mu}_j$ and $\hat{\sigma}_j$ be the sample mean and variance of $\{X_{i,j}\}_{i=1}^n$. Many different possible test statistics. Somewhat natural to consider statistics that take large values thehn some of $\hat{\mu}_j$'s are large. In this paper focus on statistic that takes large values when at least one of $\hat{\mu}_j$ are large.

In specific, focus on the following test statistic:

$$T = \max_{1 \le j \le p} \frac{\sqrt{n}\hat{\mu}_j}{\hat{\sigma}_j} \tag{5}$$

Large values of T indicate a likely violation of H_0, so it is natural to consider tests of the form

$$T > c \implies \text{reject } H_0$$

where c is appropriately chosen so that the test approximately has size $\alpha \in (0, 1)$. Consider various ways for calculating critical values and prove their validity.

2.4 Critical Values

Now move to define critical values for T such that under H_0, the probability of rejecting H_0 does not exceed size α asymptotically. Methods are ordered bu increasing computational complexity, increasing strength of required conditions, and also increasing power. Basic idea for the construction of critical values for T lies in the fact, that, under H_0:

$$T \le \max_{1 \le j \le p} \frac{\sqrt{n}(\hat{\mu}_j - \mu_j)}{\hat{\sigma}_j}$$

Consider two approaches to constructing such critical values: self-normalized and bootstrap methods. Also consider two- and three-step variants of the methods by incorporating inequality selection.

Following notation used:

$$Z_{ij} = (X_{ij} - \mu_j)/\sigma_j \quad \text{and} \quad Z_i = (Z_{i1}, \ldots, Z_{ip})^T$$

Observe that $\mathbb{E}[Z_{ij}] = 0$ and $\mathbb{E}[Z_{ij}^2] = 1$. Define

$$M_{n,k} = \max_{1 \le j \le p} \left(\mathbb{E}\left[|Z_{1,j}|^k \right] \right)^{1/k}, k = 3, 4, \quad \text{and } B_n = \left(\mathbb{E}\left[\max_{1 \le j \le p} Z_{1j}^4 \right] \right)^{1/4}$$

The dependence on n comes via the dependence of $p = p_n$ on n implicitly. By Jensen's inequality, $B_n \ge M_{n,4} \ge M_{n,3} \ge 1$. In addition, if all Z_{ij}'s are bounded a.s by a constant C, we have that $C \ge B_n$. These are useful to get a sense of various conditions on $M_{n,3}, M_{n,4}$ and B_n imposed in the theorems below.

2.4.1 Self Normalized Critical Values

One-step method: Self-normalized method considered is based on the union bound combined with moderate deviation inequality for self-normalized sums. Under H_0

$$\mathbb{P}(T > c) \le \sum_{j=1}^{p} \mathbb{P}(\sqrt{n}(\hat{\mu}_j - \mu_j)/\hat{\sigma}_j > c) \tag{6}$$

This bound seems crude when p is large. However, will exploit the self normalizing $\sqrt{n}(\hat{\mu}_j - \mu_j)/\hat{\sigma}_j$ to show that RHS of above is bounded, even if c is growing logarithmically fast with p. Using such a c will yield a test with better power properties.

For $j = 1, \ldots, p$, define

$$U_j := \sqrt{n}\mathbb{E}_n[Z_{ij}]/\sqrt{\mathbb{E}_n[Z_{ij}^2]}$$

Simple algebra yields, we see that

$$\sqrt{n}(\hat{\mu}_j - \mu_j)/\hat{\sigma}_j = U_j/\sqrt{1 - U_j^2/n}$$

where the right-hand side is increasing in U_j as long as $U_j \geq 0$. So under H_0,

$$\mathbb{P}(T > c) \leq \sum_{j=1}^{p} \mathbb{P}\left(U_j > c/\sqrt{1 + c^2/n}\right), \ c \geq 0 \tag{7}$$

Moderate deviation inequality for self-normalized sums of Jing, Shao, and Wang (2003) implies that for moderatley large $c \geq 0$,

$$\mathbb{P}\left(U_j > c/\sqrt{1 + c^2/n}\right) \approx \mathbb{P}\left(Z > x/\sqrt{1 + c^2/n}\right)$$

where $Z \sim N(0, 1)$. The above approximation holds even if Z_{ij} only have $2 + \delta$ finite moments for some $\delta > 0$. Therefore, take the critical value as

$$c^{SN}(\alpha) = \frac{\Phi^{-1}(1 - \alpha/p)}{\sqrt{1 - \Phi^{-1}(1 - \alpha/p)^2/n}} \tag{8}$$

where $\Phi(\cdot)$ is the normal cdf. We call $c^{SN}(\alpha)$ the one-step SN critical value with size α as it's derivation depends on the moderate deviation inequality for self-normalized sums. Note that

$$\Phi^{-1}(1 - \alpha/p) \sim \sqrt{\log(p/\alpha)}$$

so $c^{SN}(\alpha)$ depends on p only through $\log(p)$. Following theorem provides a non asymptotic bound on the probability that the rest statsistic T exceeds the SN critical value $c^{SN}(\alpha)$ under H_0 and shows that the bound converged to α under mild regularity conditions, validating the SN method.

Theorem 1. *(Validity of one-step SN method). Suppose that $M_{n,3}\Phi^{-1}(1 - \alpha/p) \leq n^{1/6}$. Then under H_0,*

$$\mathbb{P}(T > c^{SN}(\alpha)) \leq \alpha\left[1 + Kn^{-1/2}M_{n,3}^3\left\{1 + \Phi^{-1}(1 - \alpha/p)\right\}^3\right]$$

where K is a universal constant. Hence, if there exists constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that

$$M_{n,3}^3 \log^{3/2}(p/\alpha) \leq C_1 n^{1/2 - c_1} \tag{9}$$

then there exists a positive constant C depending only on C_1 such that under H_0,

$$\mathbb{P}(T > c^{SN}(\alpha)) \leq \alpha + Cn_{-c_1} \tag{10}$$

Moreover, this bound holds uniformly over all distributions $\mathcal{L}_X$ satisfying the moment conditions as well as the above requirement (9). In addition, if (9) holds, all components of X_1 are independent, $\mu_j = 0$ for all $1 \leq j \leq p$ and $p = p_n \to \infty$, then

$$\mathbb{P}(T > c^{SN}(\alpha)) \to 1 - e^{-\alpha}$$

I think the last bit is just to show that the test is approximately non-conservative.

Two-step method: Now move to combine the SN method with inequality selection. Motivation for doing this is that when $\mu_j < 0$ for some $j = 1, \ldots, p$ the inequality in (6) becomes strict. So, when there are many j for which μ_j are negative and large in absolute value, the resulting rest with one-step SN critical values would tend to be unnecessarily conservative. So, in order to improve the power of the test, it is better to exclude j for which μ_j are below some (negative) threshold when computing critical values.

Formally, let $0 < \beta_n < \alpha/2$ be some constant. For generality, allow β_n to depend on n. In particular, we

allow $\beta_n = o(1)$. Let $c^{SN}(\beta_n)$ be the SN critical value with size β_n and define the set $\hat{J}_{SN} \subset \{1, \ldots, p\}$ by

$$\hat{J}_{SN} := \left\{ j \in \{1, \ldots, p\} : \sqrt{n}\hat{\mu}_j / \hat{\sigma}_j > -2c^{SN}(\beta_n) \right\} \tag{11}$$

Let $\hat{k} = |\hat{J}_{SN}|$. Then, the two step SN critical value is defined by

$$c^{SN,2S}(\alpha) = \begin{cases} \frac{\Phi^{-1}(1-(\alpha-2\beta_n)/\hat{k})}{\sqrt{1-\Phi^{-1}(1-(\alpha-2\beta_n)/\hat{k})}}, & \text{if } \hat{k} \geq 1 \\ 0, & \text{if } \hat{k} = 0 \end{cases} \tag{12}$$

Then paper claims the following theorem

Theorem 2. *Suppose there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$M_{n,3}^3 \log^{3/2}\left(\frac{p}{\beta_n \wedge (\alpha - 2\beta_n)}\right) \leq C_1 n^{1/2-c_1}$$
$$\text{and } B_n^2 \log^2(p/\beta_n) \leq C_1 n^{1/2-c_1}$$

Then there exist positive constants c, C depending only on α, c_1, C_1 such that under H_0,

$$\mathbb{P}(T > c^{SN,2S}(\alpha)) \leq \alpha + Cn^{-c} \tag{13}$$

Moreover, this bound holds uniformly over all distribution $\mathcal{L}_X$ satisfying (6) and the above condition. In addition, if all components of X_1 are independent, $\mu_j = 0$ and $p = p_n \to \infty$ while $\beta_n \to 0$ then

$$\mathbb{P}(T > c^{SN,2S}(\alpha)) \to 1 - e^{-\alpha}$$

2.4.2 Bootstrap Methods

Section considers Multiplier Bootstrap and Empitical Bootstrap methods. These methods are computationally harder but they lead to less conservative tests.

One-Step Method First consider the one-step method (without moment selection). In order to make the test have size α, it is enough to choose the critical value as a bound on the $(1-\alpha)$ quantile of the distribution of

$$\max_{1 \leq j \leq p} \sqrt{n}(\hat{\mu} - \mu_j)/\hat{\sigma}_j$$

The self normalizing method finds such a bound using the union bound and moderate deviation inequality for self-normalized sums. However, SN method may be conservative as it ignores correlation between the coordinates in X_i.

Alternatively, we consider a Gaussian approximation. Under suitable regularity conditions

$$\max_{i \leq j \leq p} \sqrt{n}(\hat{\mu} - \mu_j)/\hat{\sigma}_j \approx \max_{1 \leq j \leq p} \sqrt{n}(\hat{\mu}_j - \mu_j)/\sigma_j = \max_{1 \leq j \leq p} \sqrt{n}\mathbb{E}_n[Z_{ij}]$$

where $Z_i = (Z_{i1}, \ldots, Z_{ip})^T$ are defined above ($Z_j = (X_j - \mu_j)/\sigma_j$). When p is fixed, the central limit theorem shows that, as $n \to \infty$,

$$\sqrt{n}\mathbb{E}_n[Z_i] \rightsquigarrow Y, \text{ with } Y = (Y_1, \ldots, Y_p)^Y \sim N(0, \mathbb{E}[Z_1 Z_1^T])$$

By the continuous mapping theorem, this gives us that

$$\max_{1 \leq j \leq p} \sqrt{n}\mathbb{E}_n[Z_{ij}] \rightsquigarrow \max_{1 \leq j \leq p} Y_j$$

so we can take the critical value to be the $(1-\alpha)$ wuantile of $\max_{1 \leq j \leq p} Y_j$. This theory does not cover when

p grows with n. Different tools should be used to derive an appropriate critical value for the test. A possible approach is to use a Berry-Esseen theorem that provides a suitable non-asymptotic bound between the distributions of $\sqrt{n}\mathbb{E}_n[Z_i]$ and Y. However, such Berry Esseen bounds require p to be small in comparasion with n in order to garuntee that the distribution of $\sqrt{n}\mathbb{E}_n Z_i$ is similar to that of Y. This approach builds on the work of (Chernozhukov, Chetverikov, and Kato, 2013, 2017) to show that, under some mild regularity conditions, the distribution of $\max_{1\leq j\leq p}\sqrt{n}\mathbb{E}_n[Z_{ij}]$ can be approximated by that of $\max_{1\leq j\leq p}$ in the sense of Kolmogrov distance even when p is larger or much larger than n.

Still, the distribution of $\max_{1\leq j\leq p} Y_j$ is typically unknown because the covariance structure of Y is unknown. So we will approximate the distribution of $\max_{1\leq j\leq p} Y_j$ by one of the following two bootstrap procedures:

Algorithm (Multiplier bootstrap)

1. Generate independent standard normal variables $\epsilon_1,\dots,\epsilon_n$ independent of the data
2. Construct the multiplier bootstrap test statistic

$$W^{MB} = \max_{1\leq j\leq p} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij}-\hat{\mu}_j)]}{\hat{\sigma}_j} \tag{14}$$

3. Calculate $c^{MB}(\alpha)$ as the conditional $(1-\alpha)$-quantile of W^{MB} given X_1^n

Algorithm (Empirical bootstrap)

1. Generate a bootstrap sample $X_1^*,\dots X_n^*$
2. Construct the empirical bootstrap test statistic

$$W^{EB} = \max_{1\leq j\leq p} \frac{\sqrt{n}\mathbb{E}_n[X_{ij}^*-\hat{\mu}_j]}{\hat{\sigma}_j} \tag{15}$$

3. Calculate $c^{EB}(\alpha)$ as the contional $(1-\alpha)$ quantile of W^{EB} given X_1^n.

We call these the one step multipler bootstrap and empirical bootsrap critical values, respectively, with size α. Can be computed with any precision using simulation.

Intuitively it is expected that the multiplier bootstrap works well since, conditional on the data, the vector

$$\left(\frac{\sqrt{n}\mathbb{E}[\epsilon_i(x_{ij}-\hat{\mu}_j)]}{\sigma_j}\right)_{1\leq j\leq p}$$

has the centered normal distribution with covariance matrix

$$\mathbb{E}_n\left[\frac{(X_{ij}-\hat{\mu}_j)}{\hat{\sigma}_j}\frac{(X_{ik}-\hat{\mu}_k)}{\hat{\sigma}_k}\right], 1\leq j,k\leq p \tag{16}$$

which should be close to the covariance matrix of the vector Y. Indeed by Theorem 2 in Chenozhukov, Chetverikov, and Kato (2015), the primary factor for the bound on the Kolmogorov [1] distance between the conditional distribution of W and the distribution of $\max_{1\leq j\leq p} Y_j$ is

$$\max_{1\leq j,k\leq p}\left|\mathbb{E}_n\left[\frac{(X_{ij}-\hat{\mu}_j)}{\hat{\sigma}_j}\frac{(X_{ik}-\hat{\mu}_k)}{\hat{\sigma}_k}\right]-\mathbb{E}[Z_{1j}Z_{1k}]\right|$$

which is shown to be small even when $p\gg n$ (under suitable conditions).

Following theorem establishes validity of the MB and EB critical values.

[1] The Kolmogorov Distance is defined as, for two pr. measures μ,ν on $\mathbb{R}$, $\text{Kolm}(\mu,\nu) := \sup_{x\in\mathbb{R}}\left|\mu\left((-\infty,x]\right)-\nu\left((-\infty,x]\right)\right|$

Theorem 3 (Validity of one-step MB and EB methods)**.** *Let $c^B(\alpha)$ stand for either $c^{MB}(\alpha)$ or $c^{EB}(\alpha)$. Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$(M_{n,3}^3 \vee M_{n,4}^2 \vee B_n)^2 \log^{7/2}(pn) \leq C_1 n^{1/2-c_1} \tag{17}$$

Then there exist positive constants c, C depending only on c_1, C_1 such that, under H_0,

$$\mathbb{P}(T < c^B(\alpha)) \leq \alpha + Cn^{-c} \tag{18}$$

In addition, if $\mu_j = 0$ for all j, then

$$\left|\mathbb{P}(T > c^B(\alpha)) - \alpha\right| \leq Cn^{-c} \tag{19}$$

Moreover both bounds hold uniformly over all distributions L_X satisfying the conditions (4) and (17).

Leave analysis of more general exchangeable weighted bootstraps in the high dimensional setting for future works. Also observe that the condition (17) required fof the validity of the one-step MB/EB methods is stronger than qhat is required for validity of the two-step SN method.

Two-step Methods Now consider combining bootstrap methods with inequality selection. To describe, let $0 < \beta_n < \alpha/2$ be some constant. As before, β_n can depend on n. Let $c^{MB}(\beta_n)$ and $c^{EB}(\beta_n)$ be one-step MB and EB critical values with size β_n, repsectively. Define the sets $\hat{J}_{MB}$ and $\hat{J}_{EB}$ by

$$\hat{J}_B := \left\{ j \in \{1, \ldots, p\} : \sqrt{n}\hat{\mu}_j/\hat{\sigma}_j > -2c^B(\beta_n) \right\}$$

Then, the two-step MB and EB critical values $c^{MB,2S}(\alpha)$ and $c^{EB,2S}(\alpha)$ are defined by the following procedures

Algorithm (Multiplier bootstrap with inequality selection).

1. Generate independent standard normal random variables $\epsilon_1, \ldots, \epsilon_n$ independent of the data X_1^n.
2. Construct the multiplier bootstrap test statistic

$$W_{\hat{J}_{MB}} = \begin{cases} \max_{j \in \hat{J}_{MB}} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_n(X_{ij} - \hat{\mu}_j)]}{\hat{\sigma}_j} & \text{if } \hat{J}_{MB} \text{ is not empty} \\ 0 & \text{otherwise} \end{cases}$$

3. Calculate $c^{MB,2S}$ as the conditional $(1 - \alpha + 2\beta_n)$-quantile of $W_{\hat{J}_{MB}}$ given the data

Algorithm (Empirical bootstrap with inequality selection).

1. Generate a bootstrap sample $X_1^*, \ldots, X_n^*$ as i.i.d draws from the empirical distribution of $X_1^n = \{X_1, \ldots, X_n\}$.
2. Construct the empirical bootstrap test statistic

$$W_{\hat{J}_{EB}} = \begin{cases} \max_{j \in \hat{J}_{EB}} \frac{\sqrt{n}\mathbb{E}_n[X_{ij}^* - \hat{\mu}_j]}{\hat{\sigma}_j} & \text{if } \hat{J}_{EB} \text{ is not empty} \\ 0 & \text{otherwise} \end{cases}$$

3. Calculate $c^{EB,2S}(\alpha)$ as the conditional $(1 - \alpha + 2\beta_n)$-quantile of $W_{\hat{J}_{EB}}$ given the data

Theorem 4 (Validity of two-step MB and EB methods)**.** *Let $c^{B,2S}(\alpha)$ stand for either $c^{MB,2S}(\alpha)$ or $c^{EB,2S}(\alpha)$. Suppose that the asusmption of Theorem 3 is satisfied. Moreover, suppose that $\log(1/\beta_n) \leq C_1 \log n$. Then there exist positive constants c, C depending only on c_1, C_1 such that under H_0,*

$$\mathbb{P}(T > c^{B,2S}(\alpha)) \leq \alpha + Cn^{-c}$$

In addition, if $\mu_j = 0$ for all $1 \leq j \leq p$, then

$$\mathbb{P}(T > c^{B,2S}(\alpha)) \geq \alpha - 3\beta_n - Cn^{-c}$$

so that under an extra assumption that $\beta \leq C_1 n^{-c_1}$ [a]

$$\left|\mathbb{P}(T > c^{B,2S}(\alpha)) - \alpha\right| \leq Cn^{-c}$$

Moreover all these bounds hold uniformly over all distributions L_X satisfying (4) and (17)

[a] which is to say β_n goes to 0 reasonable fast

It is sort of interesting to note that all these theorems are "non-asymptotic" in the sense that if the conditions hold then these inequalities "really" hold.

2.4.3 Hybrid Methods

Have considered one-step SN, MB, and EB methods and their two-step variatns. In fact, can also consider hybrids of these methods. For example, can use the SN method for inquality selection and then apply the MB or EB method for the selected inequalities, which is computationally more tractable. Notate this as the HB method. Formally, let $0 < \beta_n < \alpha/2$ be some constants and recall the set $\hat{J}_{SN} \subset \{1, \ldots, p\}$ defined above. Then the hybrid MB critical value, $c^{MB,H}(\alpha)$ is defined by the following procudure:

Algorithm (Multiplier Bootstrap Hybrid method).

1. Generate independent standard normal random variables $\epsilon_1, \ldots, \epsilon_n$ independent of the data X_1^n.
2. Construct the bootstrap test statistic:

$$W_{\hat{J}_{SN}} = \begin{cases} \max_{j \in \hat{J}_{SN}} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_n(X_{ij} - \hat{\mu}_j)]}{\hat{\sigma}_j} & \text{if } \hat{J}_{SN} \text{ is not empty} \\ 0 & \text{otherwise} \end{cases}$$

3. Calvculate $c^{MB,H}(\alpha)$ as the contional $(1 - \alpha + 2\beta_n)$-quantile of $W_{\hat{J}_{SN}}$ given the data.

This can be equibalently defined for the empirical bootstrap.

Theorem 5 (Validity of hybrid two-step methods)**.** *Let $c^{MB,H}$ stand either for $c^{MB,H}(\alpha)$ or $c^{MB,H}(\alpha)$. Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that (17) is verified. Moreover, suppose that $\log(1/\beta_n) \leq C_1 \log n$. Then all the conclusions of Theorem 4 hold wilth $c^{B,MS}(\alpha)$ replaced by $c^{B,H}(\alpha)$.*

2.4.4 Three-step method

In empirical studies based on moment inqualites one generally has inequalities of the form

$$\mathbb{E}[g_j(\xi, \theta)] \leq 0 \qquad \text{for all } j = 1, \ldots, p \tag{20}$$

where ξ is a vector of r.v's from a distribution denoted $\mathcal{L}_\xi$, $\theta = (\theta_1, \ldots, \theta_r)^T$ is a vector of parameters in $\mathbb{R}^r$ and $g_1, \ldots, g_p$ a set of (known) functions. In these studies, inequalities (1) and (2) arise when one tests the null hypothesis $\theta = \theta_0$ against the alternative $\theta \neq \theta_0$ on the i.i.d data $\xi_1, \ldots \xi_n$ by setting $X_{ij} := g_j(\xi_i, \theta_0)$ and $\mu_j := \mathbb{E}[X_{1j}]$. So far, have shown how to increase the poser of such tests by employing inequality selection procedures that allow the researcher to drop uninformative inequalities. In this subsection, combine this

selection procedure with another procedure suitable for the model (20) by dropping *weakly informative* inequalities, that is inequalities j with the function $\theta \mapsto \mathbb{E}[g_j(\xi, \theta)]$ being flat or nearly flat around $\theta = \theta_0$.

When the tested value θ_0 is close to some θ satisfying (20), such inequalities can only provide a weak signal of violation of the hypothesis $\theta = \theta_0$ in the sense that they have $\mu_j \approx 0$ and so it is useful to drop them. For brevity, only consider weakly informative inequality selection based on the MB and EB methods and note that similar results can be obtained for the self-normalized method. Also only consider the case where the function $\theta \mapsto g_j(\xi, \theta)$ are almost everywhere continuously differentiable and leave the extension to non-differentiable functions to future work.

Start with the necessary notation. Let $\xi_1, \ldots, \xi_n$ be a sample of observations from the distribution of ξ. Suppose that we are interested in testing the null hypothesis and alternative hypothesis

$$H_0 : \mathbb{E}[g_j(\xi, \theta_0)] \leq 0 \quad \text{for all } j = 1, \ldots, p$$
$$H_a : \mathbb{E}[g_j(\xi, \theta_0)] > 0 \quad \text{for some } j = 1, \ldots, p$$

where θ_0 is some value of the parameter θ. Define

$$m_j(\xi, \theta) := \big(m_{j1}(\xi, \theta), \ldots, m_{jr}(\xi, \theta)\big)^T$$
$$:= \left(\frac{\partial g_j(\xi, \theta)}{\partial \theta_1}, \ldots, \frac{\partial g_j(\xi, \theta)}{\partial \theta_r}\right)^T$$

Further, let $X_{ij} := g_j(\xi_i, \theta_0), \mu_j := \mathbb{E}[X_{1j}], \sigma_j := \big(\mathrm{Var}(X_{1j})\big)^{1/2}, V_{ijl} := m_j l(\xi_i, \theta_0), \mu_{jl}^B = \mathbb{E}[V_{1jl}]$, and $\sigma_{jl}^V := \big(\mathrm{Var}(V_{1jl})\big)^{1/2}$. Assume that

$$\mathbb{E}[X_{1,j}^2] < \infty, \sigma_j > 0, j = 1, \ldots, p \tag{21}$$
$$\mathbb{E}[X_{1,j,l}^2] < \infty, \sigma_{jl}^V > 0, j = 1, \ldots, p, l = 1, \ldots, r \tag{22}$$

In addition, let $\hat{\mu}_j = \mathbb{E}_n[X_{ij}]$ and $\hat{\sigma}_j = \left(\mathbb{E}\left[(X_{ij} - \hat{\mu}_j)^2\right]\right)^{1/2}$ be estimators of μ_j and σ_j, respectively. Similarly let $\hat{\mu}_{jl}^V = \mathbb{E}_n[V_{ijl}]$ and $\hat{\sigma}_{jl}^V = \left(\mathbb{E}\left[(V_{ij} - \hat{\mu}_{jl}^V)^2\right]\right)^{1/2}$ be esttimators of μ_{jl}^V. The inequality selection derived is similar to the bootstrap methods described in Section 4

Algorithm(Multiplier bootstrap for gradient statistic).

1. Generate independent standard normal variables $\epsilon_1, \ldots, \epsilon_n$ independent of the data.
2. Construct the multiplier bootstrap gradient statistic

$$W_{MB}^V = \max_{j,l} \frac{\sqrt{n}|\mathbb{E}_n[\epsilon_i(V_{ijl} - \hat{\mu}_{jl}^V)]|}{\hat{\sigma}_{jl}^V} \tag{23}$$

3. For $\gamma \in (0, 1)$, calculate $c^{MB,V}(\gamma)$ as the conditional $(1 - \gamma)$ quantile of W_{MB}^V given the data.

Algorithm(Empirical bootstrap for gradient statistic).

1. Generate a bootstrap sample $V_1^*, \ldots, V_n^*$ as i.i.d draws from the data
2. Construct the empirical bootstrap gradient statistic

$$W_{EB}^V = \max_{j,l} \frac{\sqrt{n}|\mathbb{E}_n[V_{ijl}^* - \hat{\mu}_{jl}^V]|}{\hat{\sigma}_{jl}^V} \tag{24}$$

3. For $\gamma \in (0, 1)$, calculate $c^{EB,V}(\gamma)$ as the conditional $(1 - \gamma)$ quantile of W_{EB}^V given the data.

For $c_2, C_2 > 0$, let φ_n be a sequence of constants satisfying $\varphi_n \log n \geq c_2$ and let β_n be a sequence of constants satisfying $0 < \beta_n < \alpha/4$ and $\log\left(1/(\beta_n - \varphi_n)\right) \leq C_2 \log n$ where α is the nominal level of the test. Define three estimated sets of inequalities

$$\hat{J}_B := \left\{ j \in \{1, \ldots, p\} : \sqrt{n}\hat{\mu}_j/\hat{\sigma}_j > -2c^B(\beta_n) \right\}$$
$$\hat{J}'_B := \left\{ j \in \{1, \ldots, p\} : \sqrt{n}|\hat{\mu}^V_{jl}/\hat{\sigma}^V_{jl}| > 3c^{B,V}(\beta_n - \phi_n) \text{ for some } l = 1, \ldots, r \right\}$$
$$\hat{J}''_B := \left\{ j \in \{1, \ldots, p\} : \sqrt{n}|\hat{\mu}^V_{jl}/\hat{\sigma}^V_{jl}| > c^{B,V}(\beta_n + \phi_n) \text{ for some } l = 1, \ldots, r \right\}$$

where B stands for either MB or EB.

The derived weakly informative inequality selection procedure requires that both the test statistic and the critical value depend on the estimated sets of inequalities. Let T^B and $c^{B,3S}$ denote the test statistic and the critical value. If the set $\hat{J}'_B$ is empty, set the test statistic and critical value $T^B = c^{B,3S} = 0$. Otherwise, define the test statistic

$$T^B = \max_{j \in \hat{J}'_B} \frac{\sqrt{n}\hat{\mu}_j}{\hat{\sigma}_j}$$

and define the three-step MB/EB critical values $c^{B,3S}(\alpha)$ for the test by the same bootstrap procedures as those for $c^{B,2S}(\alpha)$ with $\hat{J}_B$ replaced by $\hat{J}' \cap \hat{J}''_B$ and also $2\beta_n$ replaced by $4\beta_n$. That is $c^{B,3S}(\alpha)$ is the conditional $(1 - \alpha + 4\beta_n)$-quantile of $W_{\hat{J}'_B \cap \hat{J}''_B}$ given the data.

Stating the main results of this section requires the following notation. Let

$$Z^V_{ijl} := (V_{ijl} - \mu^V_{ijl})/\sigma^V_{jl} \quad \text{and} \quad M^V_{n,k} := \max_{j,l} \left(\mathbb{E}[|Z^V_{ijl}|^k] \right)^{1/k} \quad \text{and} \quad B^V_n := \left(\mathbb{E}[\max_{j,l}(Z^V_{ijl})^4 \right)^{1/4}$$

Theorem 6 (Validity of three-step MB and EB methods)**.** [a] *Let T^B and $c^{B,3S}(\alpha)$ stand for T^{MB} and $c^{MB,3S}(\alpha)$ or for T^{EB} and $c^{EB,3S}(\alpha)$. Suppose there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$\left(M^3_{n,3} \vee M^2_{n,4} \vee B_n)^2 \log^{7/2}(pn) \right) \leq C_1 n^{1/2 - c_1} \tag{25}$$

and

$$\left((M^V_{n,3})^3 \vee (M^V_{n,4})^2 \vee (B^V_n)^2 \log^{7/2}(pn) \right) \leq C_1 n^{1/2 - c_1} \tag{26}$$

Moreover, suppose that $\log(1/(\beta_n - \phi_n)) \leq C_2 \log n$ and $\phi_n \log n \geq c_2$ for some constants $c_2, C_2 > 0$. Then there exist positive constants c, C depending only on c_1, C_1, c_2, C_2 such that, under H_0,

$$\mathbb{P}(T^B > c^{B,3S}(\alpha)) \leq \alpha + Cn^{-c}$$

and this bound holds uniformly over all distributions L_ξ satisfying (21),(22),(25), and (26).

[a] If I understand correctly, this method only selects out weakly uninformative inequalites based on the gradient, not necessarily inequalities with say, $\hat{\mu}_j \ll 0$.

2.5 Power

Consider the same general setup as described in the intruction and assume that (4) holds. Pick any $\alpha \in (0, 1/2)$ and consider the test of the form

$$T > \hat{c}(\alpha) \implies \text{reject } H_0$$

Where $\hat{c}(\alpha)$ is equal to $c^{SN}(\alpha), c^{SN,2S}(\alpha), c^{MB}(\alpha), c^{MB,2S}(\alpha), c^{EB}(\alpha), c^{EB,2S}(\alpha), c^{MB,H}(\alpha)$, or $c^{EB,H}(\alpha)$.[1] The following result holds:

[1] Importantly, the three step procedure is not included here.

Theorem 7 (Rate of uniform consistency)**.** *Suppose there exist constants* $0 < c_1 < 1/2$ *and* $C_1 > 0$ *such that*

$$M_{n,4}^2 \log^{1/2} p \leq C_1 n^{1/2-c_1} \text{ and } \log^{3/2} p \leq C_1 n \tag{27}$$

In addition, suppose that $\inf_{n\geq 1}(\alpha - 2\beta_n) \geq c_1\alpha$ *whenever inequality selection is used. Then there exist constants* $c, C > 0$ *depending only on* α, c_1, C_1 *such that for every* $\epsilon \in (0,1)$*, whenever*[a]

$$\max_{1\leq j\leq p} \mu_j/\sigma_j \geq (1 + \epsilon + C\log^{-1/2} p)\sqrt{\frac{2\log(p/\alpha)}{n}}$$

we have

$$\mathbb{P}(T > \hat{c}(\alpha)) \geq 1 - \frac{C}{\epsilon^2 \log(p/\alpha)} - Cn^{-c}$$

Therefore, when $p = p_n \to \infty$*, for any sequence* ϵ_n *satisfying* $\epsilon_n \to 0$ *and* $\epsilon_n\sqrt{\log p_n} \to \infty$*, as* $n \to \infty$*, we have*

$$\inf_{\mu\in\mathcal{B}_n} \mathbb{P}_\mu(T > \hat{c}(\alpha)) \geq 1 - o(1) \tag{28}$$

where

$$\mathcal{B}_n = \left\{\mu = (\mu_1, \ldots, \mu_p) : \max_{1\leq j\leq p} \mu_j/\sigma_j \geq \bar{r}_n(1+\epsilon_n)\sqrt{2\log(p_n)/n}\right\}$$

and P_μ *denotes the probability measure for the distribution* $\mathcal{L}_X$ *having mean* μ*. Moreover, the above asymptotic result (28) holds uniformly with respect to any sequnce of distributions* $\mathcal{L}_X$ *satisfying (4) and (27).*

[a] Seems similar to LASSO condition

Theorem 7 shows that tests are uniformly consistent against all alternatives excluding those in a small neighborhood of alternatives that are too close to the null. Size of this neighborhood is shrikning at a fast rate. They show in a working paper that no test can be consistent against all alternatives whose distance from the null converges to zero faster than $\sqrt{(\log p_n)/n}$ and that the tests above are minimax optimal.

2.6 Monte Carlo Experiments

Results are given, I would check document. This likely concludes my notes on this topic.

3 Set Identification in Models with Multiple Equilibria; *Alfred Galichon, Marc Henry (ReStud, 2011)*

Galichon and Henry (2011) proposes a computationally feasible way of deriving the identified features of a model with multiple equilibria in pure or mixed strategies. It can be found from the ReStud website here.

3.1 Introduction

Empirical study of game theoretic models generally complicated by the presence of multiple equilibria. The existence of multiple equilibria generally leads to a failure of identification of the structural parameters governing the model.

- Berry and Tamer (2006) and Ackerberg et al. (2007) give an account of the various ways this issue is approached in the literature.
- Andrews, Berry, and Jia (2003), Ciliberto and Tamer (2009) consider some partial identification approach. Identification approach is not sharp.

Paper proposes a computationally feasible way of recovering the identified. Note a generalized likelihood implied by a model with multiple equilibria can be represented by a non-additive set function called a *Choquet Capacity*. Give a formal definition of an equilibrium selection mechanism and call such a mechanism compatible with the data if the likelihood of the model augmented with such a mechanism is equal to the probabilities observed in the data. THe identified feature of the model is then the set of parameter values s.t there exists an equilibrium selection mechanism compatible with the data.

The computational burden remains high in situations with a large number of observable outcomes since the number of inequalities to be checked is equal to the number of subsets of the set of observable outcomes.

3.2 Identified Features of Models with Multiple Equilibria

First go over framework and general results in the case where only equilibria in pure strategies are considered. Section 1.2 specializes and illustrates them on leading examples of participation games.

3.2.1 Identified Parameter Sets in General Models with Multiple Equilibria

General framework is that of Jovanovic (1989). Applies to the empirical analysis of normal form games, where only equilibra in pure strategies are considered. Consider three types of economic variables.

- Outcome variables, Y
- Exogenous explanatory variables X
- Random shocks (or latent variables) ϵ

Outcome variables and latent variables are assumed to belong to complete and separable metric spaces. Economic model consists of a set of restrictions on the joint behavior of the variables listed above. Restrictions may be induced by assumptions of rational agents, and they generally depend on a set of unknown structural parameters, θ.

Without loss of generality, model may be formalized as a measurable correspondence (defined below) between the latent variables ϵ and the outcome variables Y, indexed by the exogenous variables X and the vector of parameters θ. This correspondence is called G and write $Y \in G(\epsilon|X;\theta)$ to indicate admissible values of Y given ϵ, X, θ. The econometrician is assumed to have access to a sample of i.i.d vectors (Y, X) and the problem that is considered is estimating the vector of parameters θ. The latent variables ϵ is supposed to be distributed according to a parametric distribution $\nu(\cdot|X;\theta)$. Assumptions are collected below:

Assumption 1. An independent and identically distributed sample of copies of the random vector (Y, X) is available. The observable outcomes Y conditionally distributed according to the probability distribution $P(\cdot|X)$ on $\mathcal{Y}$, a Polish space[a] endowed with its Borel σ-algebra of subsets $\mathcal{B}$ are related to the unobservable variables ϵ according to the model $Y \in G(\epsilon|X;\theta)$. Here, θ belongs to an open subset Θ of $\mathbb{R}^{d_\theta}$, ϵ is distributed according to the probability measure $\nu(\cdot|X;\theta)$ on $\mathcal{U}$[b], and G is a measurable correspondence[c] for almost all X and for all $\theta \in \Theta$. Finally, the variables (Y, X, ϵ) are defined on the same underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

[a] A Polish space is a complete and separable metric space. A complete metric space is one where every Cauchy sequence converges to a point in the space and a separable metric space is one with a countable, dense, subset. A Cauchy sequence, $(a_n)_{n\in\mathbb{N}}$ is one such that for every $\epsilon > 0, \exists N_\epsilon \in \mathbb{N}$ s.t $\forall m, n \geq N_\epsilon, d(a_m, a_n) < \epsilon$. Completeness ensures that the space is "rich enough." For example, the reals are complete but the rationals are not.

[b] Also a Polish space

[c] A measurable correspondence is such that for all open subsets $A \subseteq \mathcal{Y}$, $G^{-1}(A|X;\theta) := \{\epsilon \in \mathcal{U} : G(\epsilon|X;\theta) \cap A \neq 0\}$ is measurable. A note: technically A should be measurable, but since we are dealing with Borel σ-algebra's this is an equivalent definition. A measurable correspondence is also called a random correspondence or a random set.

Example 1. To illustrate, consider a simple game proposed by Jovanovic (1989). Consider two firms with profit functions $\Pi_1(Y_1, Y_2, \epsilon_1, \epsilon_2; \theta) = (\theta Y_2 - \epsilon_2)Y_1$ and $\Pi_2(Y_1, Y_2, \epsilon, \epsilon_2; \theta) = (\theta Y_1 - \epsilon_1)Y_2$ where $Y_i \in \{0, 1\}$ is firm i's action and $\epsilon = (\epsilon_1, \epsilon_2)$ are exogenous cots. The firms know their costs, the analyst only knows that ϵ is uniformly distributed on $[0, 1]^2$ and that the structural parameter θ is in $(0, 1]$. There are two pure strategy Nash equilibria. The first is $Y_1 = Y_2 = 0$ for all $\epsilon \in [0, 1]^2$. The second is $Y_1 = Y_2 = 1$ for all $\epsilon \in [0, \theta]^2$ and $Y_1 = Y_2 = 0$ otherwise. Hence the model is described be the correspondence $G(\epsilon; \theta) = \{(0, 0), (1, 1)\}$ for all $\epsilon \in [0, \theta]^2$ and $G(\epsilon; \theta) = \{0, 0\}$ otherwise.

To conduct inference on the parameter vector θ, one first needs to determine the identified features of the model. Since G may be multi-valued due to the presence of multiple equilibria, the outcomes may not be uniquely determined by the latent variable. In such cases, the generalized likelihood of an outcome falling in the subset A of $\mathcal{Y}$ predicted by the model is $\mathcal{L}(A|X;\theta) = \nu(G^{-1}(A|X;\theta)|X;\theta)$. Because of multiple equilibria, this generalized likelihood may sum to more than one, as we may have $A \cap B = \emptyset$ and yet $G^{-1}(A|X;\theta) \cap G^{-1}(B|X;\theta) \neq 0$[1] so that for these sets A, B such that $A \cap B = \emptyset$, $\mathcal{L}(A \cup B|X;\theta) < L(A|X;\theta) + L(B|X;\theta)$. The set function $A \mapsto \mathcal{L}(A|X;\theta) = \nu(G^{-1}(A|X;\theta)|X;\theta)$ is generally not additive and is called a *Choquet capacity*[2]. This non-additivity of the model likelihood is well documented.

Definition 1 (Choquet capacity)**.** A Choquet capacity $\mathcal{L}$ on a finite set $\mathcal{Y}$ is a set function $\mathcal{L} : A \subset \mathcal{Y} \mapsto [0, 1]$ which is

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

This is like a probability measure but without additivity. In the example above, $\nu(\cdot|X;\theta)$ is the uniform distribution on $[0, 1]^2$ and the Choquet capacity $\nu(G^{-1})$ gives value $\nu(G^{-1}(\{0, 0\})) = \nu([0, 1]^2) = 1$. and $\nu(G^{-1}\{1, 1\}) = \nu([0, \theta]^2) = \theta^2$ to the set $\{(1, 1)\}$. Hence it is immediately apparent that the Choquet capacity of $\nu(G^{-1})$ is nonadditive.

As discussed in Jovanovic (1989) and Berry and Tamer (2006), the model with multiple equilibria can be completed with an equilibrium selection mechanism. Define an equilibrium selection mechanism as a conditional distribution $\pi_{Y|\epsilon,X;\theta}$ over equilibrium outcomes Y in the regions of multiplicity. By construction, an equilibrium selection mechanism is allowed to depend on the latent variables ϵ even after conditioning on X.

Definition 2 (Equilibrium selection mechanism)**.** An equilibrium selection mechanism is a conditional probability $\pi(\cdot|\epsilon, X;\theta)$ for Y conditional on ϵ and X such that the selected value of the outcome variable is actually an equilibrium. Formally $\pi(\cdot|\epsilon, X, \theta)$ has support contained in $G(\epsilon|X;\theta)$.

[1] Intuition: $G^{-1}(y|X;\theta)$ gives the set of ϵ values that could have generated the y value (observed outcome) conditional on X and θ and one epsilon can generate two different y values because of multiple equilibria.

[2] See Choquet, 1954. Choquet capacity also used as a generalized probability in some behavioral decision making theory.

Crucial to this is the fact that π is a *probability measure*. It should "smooth out" the non-additivity of $\nu(G^{-1})$.

The identified feature of the model is the smallest set of parameters that cannot be rejected by the data. Hence, it is the set of parameters for which one can find an equilibrium selection mechanism that completes the model and equates probabilities of outcomes predicted by the model with the probabilities obtained from the data.[3]

Definition 3 (Compatible equilibrium selection mechanism)**.** The equilibrium selection mechanism $\pi(\cdot|\epsilon, X; \theta)$ is compatible with the data if the probabilities observed in the data are equal to the probabilities predicted by the equilibrium selection mechanism. More formally, if for all measurable subsets A of $\mathcal{Y}$

$$P(A|X) = \int_{\mathcal{U}} \pi(A|\epsilon, X, \theta)\nu(d\epsilon|X; \theta)$$

Definition 4 (Identified Set)**.** The identified set (or the *sharp* identified set) is the set $\theta_I \subseteq \Theta$ such that, $\forall\theta \in \theta_I$, there exists an equilibrium selection mechanism compatible with the data.

Above definition is not operational, in the sense that it does not allow for the computation of the identified set based on the knowledge of the probabilities in the data, because π is an infinite dimensional nuisance parameter. Now set out to show how to reduce the dimensionality of the problem. Equivalent formulation of the identified set relates to the *core* of the Choquet capacity.

Definition 5 (Core of a Choquet capacity)**.** The *core* of a Choquet capacity $\mathcal{L}$ on $\mathcal{Y}$ is the collection of probability distributions Q on $\mathcal{Y}$ such that for all $A \subset \mathcal{Y}$, $Q(A) \leq \mathcal{L}(A)$. [a]

[a]Equivalently, if we consider the random set $\mathcal{L}$ as a map from $(\Omega, \mathcal{F}, \mathbb{P}) \to 2^{\mathcal{Y}}$, we can say a random variable $\gamma : \Omega \to \mathcal{Y}$ is in the core of $\mathcal{L}$ if $\gamma(\omega) \in \mathcal{L}(\omega), \forall\omega \in \Omega$. The random variable γ induces a distribution on $\mathcal{Y}$ that has the above property, and every distribution on $\mathcal{Y}$ with the above property should be induced by a random variable with this property.

In cooperative game theory, a Choquet capacity on a set $\mathcal{Y}$ is interpreted as a game, where $\mathcal{Y}$ is the set of players and $\mathcal{L}$ is the utility value or worth of a coalition $A \subseteq \mathcal{Y}$ and the core of the game $\mathcal{L}$ is the collection of allocations that cannot be improved upon by any coalition of players.

In Example 1, the core of the Choquet capacity νG^{-1} is the set of probabilities P for the observed outcomes $(0, 0)$ and $(1, 1)$ such that $P(\{(0, 0)\}) \leq \nu G^{-1}(\{(0, 0\}) = \nu([0, 1]^2) = 1$ and $P(\{(1, 1)\}) \leq \nu G^{-1}(\{(1, 1)\}) = \nu([0, \theta]^2) = \theta^2$.

Next result shows the equivalence between the existence of a compatible eqm. selection mechanism and the fact that the true distribution of the data belongs to the core of the Choquet capacity that characterizes the generalized likelihood predicted by the model.

Theorem 1. *The identified set Θ_I is the set of parameters such that the true distribution of the observable variables lies in the core of the generalized likelihood predicted by the model.*

$$\Theta_I = \left\{\theta \in \Theta : \forall A \in \mathcal{B}, P(A|X) \leq \mathcal{L}(A|X; \theta); X - a.s\right\}$$

A later theorem generalizes this to the case of mixed strategy eqm. In the example above, the identified set is the set of values for θ such that $0 \leq \mathbb{P}((Y_1, Y_2)) \leq \theta^2$.

The first thing to note from this theorem is that the problem of computing the identified set has been transformed into a finite-dimensional problem in the case where $\mathcal{Y}$ is a finite set. Indeed, in this case, the problem of computing the identified set is reduced to the problem of checking a finite number of inequalities.

However, in cases where the cardinality of $\mathcal{Y}$ is large, then the number of inequalities to be checked is $2^{|\mathcal{Y}|} - 2$ and the computational burden is only partially lifted. The rest of the paper is based on the characterization of Theorem 1.

[3]Is this somehow restrictive? I guess not, since we've placed no assumption on the selection mechanism.

3.2.2 Some illustrative examples

Examples are given in this section of Market Entry, Family Bargaining etc. For the most part, they resemble the market entry example given above in Section 2.

Family Bargaining Go over this game since it is later considered in the optimal transport section. Consider a simplified version of the bargaining model of decision regarding the long-term care of an elderly parent for a family with two children. The issue is which child will care for the parent when the parent ages or whether the parent is moved to a nursing home. The payoff to family member $i, i = 1, 2$ is represented by the sum of three terms.

The first term, V_{ij} represents the value to child i of care option j, where $j > 0$ means child j becomes the primary care giver, and $j = 0$ means the parent is moved to a nursing home. The matrix $(V_{ij})_{ij}$ is known to both children. We suppose it takes the form

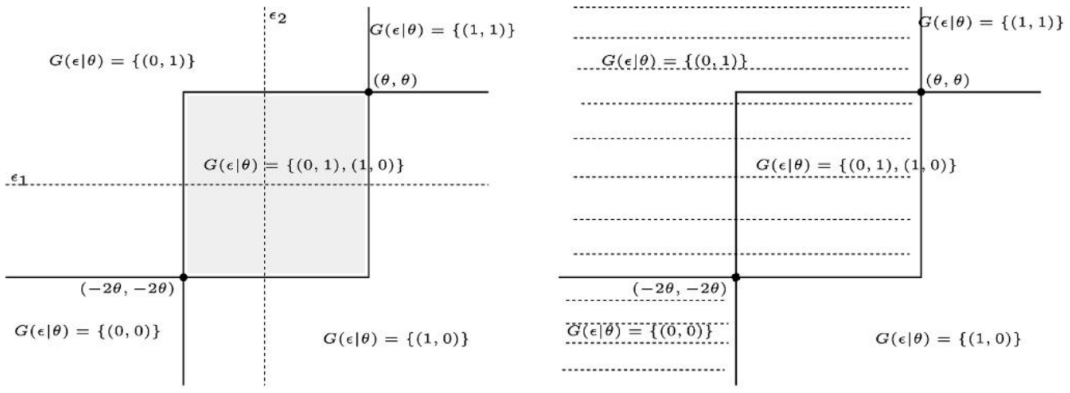
$$V = \begin{pmatrix} 0 & 2\theta & 4\theta \\ 0 & 2\theta & 4\theta \end{pmatrix}$$

$\theta > 0$ is unknown to the analyst. Both children simultaneously decide whether or not to take part in the long-term care decision. Suppose M is the set of children who participate. The option chosen is option j that maximizes the sum $\sum_{i \in M} V_{ij}$ among the available options. It is assumed that participants abide with the decision and that benefits are then shared equally amongst the children participating int he decision through a monetary transfer s_i, which is the second term in the children's payoff. Third term ϵ_i is a random benefit from participation, which is 0 for children who decide not to participate and distributed according to $\nu(\cdot|\theta)$ for children who participate[4]. All players observe the realization of ϵ, while the analyst knows only its distribution.

Equilibria correspondence, restricting analysis to only pure strategy NE is:

$$\begin{aligned}
&\{(0,0)\}\text{is a Nash Equilibrium in pure strategies iff } \epsilon_2 < -2\theta \text{ and } \epsilon_1 < -2\theta \\
&\{(1,1)\}\text{is a Nash Equilibrium in pure strategies iff } \epsilon_2 > \theta \text{ and } \epsilon_1 > \theta \\
&\{(0,1)\}\text{is a Nash Equilibrium in pure strategies iff } \epsilon_2 > -2\theta \text{ and } \epsilon_1 < \theta \\
&\{(1,0)\}\text{is a Nash Equilibrium in pure strategies iff } \epsilon_2 < \theta \text{ and } \epsilon_1 > -2\theta
\end{aligned}$$

The equilibrium correspondence $G(\epsilon|\theta)$ is represented in Figure 1(a) below.



(a) Model Correspondence (b) Inequality to check

Figure 1: Family Game [Lifted from Paper]

[4]Normalizing the utility of the outside option to 0

In the case of the family bargaining game, the set of possible outcomes is $\mathcal{Y} = \{(0,0),(0,1),(1,0),(1,1)\}$. The generalized likelihood of outcomes predicted by the model can be written as follows:

$$\begin{aligned}
\mathcal{L}(\{(0,0)\}|\theta) &= \nu(\epsilon : \epsilon_1 \leq -2\theta, \epsilon_2 \leq -2\theta|\theta) = \nu(G^{-1}((0,0)|\theta)|\theta) \\
\mathcal{L}(\{(0,1)\}|\theta) &= \nu(\epsilon : \epsilon_1 \leq \theta, \epsilon_2 \geq -2\theta|\theta) = \nu(G^{-1}((0,1)|\theta)|\theta) \\
\mathcal{L}(\{(1,0)\}|\theta) &= \nu(\epsilon : \epsilon_1 \geq -2\theta, \epsilon_2 \leq \theta|\theta) = \nu(G^{-1}((1,0)|\theta)|\theta) \\
\mathcal{L}(\{(1,1)\}|\theta) &= \nu(\epsilon : \epsilon_1 \geq \theta, \epsilon_2 \geq \theta|\theta) = \nu(G^{-1}((1,1)|\theta)|\theta)
\end{aligned}$$

anf the generalized likelihood of the remaining events can be derived as follows

$$\begin{aligned}
\mathcal{L}(\{(0,0)\} \cup A|\theta) &= \mathcal{L}(\{(0,0)\}|\theta) + \mathcal{L}(A|\theta), \quad \text{for all } A \subset \mathcal{Y}/\{(0,0)\} \\
\mathcal{L}(\{(1,1)\} \cup A|\theta) &= \mathcal{L}(\{(1,1)\}|\theta) + \mathcal{L}(A|\theta), \quad \text{for all } A \subset \mathcal{Y}/\{(1,1)\} \\
\mathcal{L}(\{(0,1),(1,0)\}|\theta) &= 1 - \mathcal{L}(\{(0,0),(1,1)\}|\theta)
\end{aligned}$$

The generalized likelihood predicted by the model is the set function $A \mapsto \mathcal{L}(A|\theta) = \nu(G^{-1}(A|\theta)|\theta)$ for $A \subset \mathcal{Y} = \{(0,0),(0,1),(1,0),(1,1)\}$. This set function is a Choquet capacity and if the support of ν is sufficiently large, the generalized likelihood sums to more than one because the region of multiple equilibria is "counted twice".

Model is completed by adding an equilibrium selection mechanism that will pick out a single equilibrium for each value of the latent variable ϵ in the region of multiplicity. As formally defined previously, an equilibrium selection mechanism is a conditional probability $\pi(\cdot|\epsilon, X, \theta)$ with support included in $G(\epsilon|X;\theta)$. It is compatible with the data if the probabilities it predicts are equal to the true probabilities of the observable variables.

In this example, for $j = 0, 1$:

$$P((i,j)|X) = \int_{\mathcal{U}} \pi((i,j)|\epsilon, X;\theta)\nu(d\epsilon|X;\theta)$$

Since the model contains no prior information, any valid probability measure equilibrium selection mechanism that generates equates predicted probabilities with observed probabilities is consistent.

It is noted that the definition of the identified region using a semi-parametric likelihood representation, with the equilibrium selection mechanism as the infinite dimensional nuisance parameter π is impracticable, so Theorem 1 is used to make it operational and compute Θ_I. So

$$\Theta_I = \{\theta \in \Theta : (\forall A \in 2^{\mathcal{Y}}; P(A|X) \leq \mathcal{L}(A|X;\theta); X - a.s\}$$

3.3 Efficient Computation of the Identified Set

Subtitle: "*Which inequalities to check and how to check them?*"

Describe three approaches to the effective computation of the identified set based on the characterization of Theorem 1. First approach is based on submodular optimization and extends readily to the case with mixed strategies. Second approach, describes in Section 2.2, relies on the highly efficient algorithms for optimal transportation problems.[1] Third approach is based off the notion of *core determining sets* and provides a dramatic reduction in the computational complexity under specific assumptions on the game under study.

3.3.1 Submodular Optimization

The first proposal to deal with the complexity of the problem of checking inequalities in Theorem 1 is a method of general validity based on the minimization of a submodular function, the discrete equivalent of

[1] I found the following ArXiv introduction to optimal transport problems here (if the link doesn't work; https://arxiv.org/abs/1009.3856)

a convex function. This is a well-known problem in combinatorial optimization and efficient algorithms are easily available off the shelf.

Definition 6 (Submodular function)**.** A set function $\mathcal{L} : \mathcal{Y} \to \mathbb{R}$ is called submodular if, for each $A, B \subset \mathcal{Y}$, we have

$$\mathcal{L}(A \cup B) + \mathcal{L}(A \cap B) \leq \mathcal{L}(A) + \mathcal{L}(B)$$

In the case that $\mathcal{L}$ is a probability measure, this holds as equality.

Submodularity for set functions is the analogue of convexity, and the problem of minimizing a submodular function is well studied. Paper now shows that checking inequalities involved in the characterization of the identified set in Theorem 1 is equivalent to the minimization of a submodular function. Theorem 1 whose that the identified set is the set of values of θ such that X-almost surely, we have the domination $\forall A \subseteq \mathcal{Y}$, $P(A|X) \leq \mathcal{L}(A|X;\theta)$. Equivalently,

$$\min_{A \subseteq \mathcal{Y}} \big(\mathcal{L}(A|X;\theta) - P(A|X)\big) \geq 0$$

First note that the function above is indeed submodular.

Lemma 1 (Submodularity of the generalized likelihood)**.** *For all $\theta \in \Theta$ and all X, the set function $\mathcal{Y}$ defined for all $A \subseteq \mathcal{Y}$ by $A \mapsto \mathcal{L}(A|X;\theta) - P(A|X)$ is submodular.*

The most efficient, generic, way to check that a convex function is everywhere non-negative and verify that the minimum is non-negative. Apply the same logic to the above. Of course, can speed this up by terminating the algorithm when a negative value is found.

Theorem 2 (Computation of the identified set)**.** *The identified set is obtained by minimization of a submodular function*

$$\theta_I = \left\{\theta \in \Theta : \min_{B \subseteq \mathcal{Y}} \big(\mathcal{L}(B|X;\theta - P(B|X)\big) = 0, X - a.s\right\}$$

As a note: I think the reason there is an "$= 0$" instead of a ≥ 0 in the statement of the identified set above is that we can always take $B = \emptyset$. More details on the procedure are given later on in Section 4. This method can be generalized to the case where equilibria in mixed strategies are considered.

The below is a special case of submodular optimization which is more efficient and applies to the case where only equilibria in pure strategies are considered.

3.3.2 Optimal Transportation Approach

When equilibria are only in pure strategies, the model generalized likelihood $\mathcal{L}$ is a very special case of submodular function since it is derived as the distribution function of a random set.

$$\mathcal{L}(A|X;\theta) = \nu(\epsilon : G(\epsilon|X;\theta) \cap A \neq \emptyset|X;\theta)$$

When mixed equilibria are considered, this improvement in efficiency is no longer available because (in general), the model generalized likelihood is no longer the distribution of a random set.[2] To describe the method, need the following notations and definitions.

Call $\mathcal{U}^*$ the set of predicted combinations of equilibrium, formally $\mathcal{U}^* = \{G(\epsilon|X;\theta); \epsilon \in \mathcal{U}\}$, remembering that $\mathcal{U}$ is the support of ϵ. Note that $\mathcal{U}^*$ is a quotient space for the correspondence G [3]. So $\mathcal{U}^*$ contains

[2] Why not? I think because the probabilities of observing an outcome now are not functions of the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, but, rather also depend on the properties of the mixed strategy equilibrium

[3] From WolframMathWorld: A quotient space, $X/\sim$ of a topological space X and a set of equivalence classes $\sim$ on X is the set of equivalence classes of points in X (under $\sim$). Open sets on $X/\sim$ can be described using the map $\pi : X \to X/\sim$ which maps each point in X to its equivalence class. A subset $W \subseteq X/\sim$ is open if $\pi^{-1}(W)$ is open. $\mathcal{U}^*$ is a quotient space for $\mathcal{U}$ using equivalence classes from the correspondence G, that is $\mathcal{U} = X/\overset{G}{\sim}$ where $\epsilon_1 \overset{G}{\sim} \epsilon_2$ if $G(\epsilon_1) = G(\epsilon_2)$

subsets of $\mathcal{Y}$ but is typically of much lower cardinality than $2^{\mathcal{Y}}$.

Further, consider the bipartite graph $\mathcal{G}(\theta, X)$ in $\mathcal{Y} \times \mathcal{U}^*$. The edges are defined as $(y, u) \in E(\mathcal{G})$ if $y \in u$. Each vertex $u \in \mathcal{Y}$ has weight $P(y|X)$ and each vertex $y \in \mathcal{U}^*$ has weight $\nu(\{\epsilon : G(\epsilon|X;\theta) = u|X\})$. Finally, call $Q(\cdot|X;\theta)$ the probabilities $Q(u|X;\theta) = \nu\{G^{-1}(u)|X;\theta)\}$ (so t at $Q(u|X;\theta)$ is the weight attached to vertex $u \in \mathcal{Y}$.)

To summarize:

1. $\mathcal{Y}$ is the set of all possible (observed) outcomes
2. $G(\epsilon|X;\theta)$ is the (measurable) *equilibrium correspondence* from the underlying probability space onto $\mathcal{Y}$. For any open subset $A \subset \mathcal{Y}$,
$$G^{-1}(A|X;\theta) := \{\epsilon \in \mathcal{U} : G(\epsilon|X;\theta) \cap A \neq \emptyset\}$$
is measurable
3. ϵ is distributed according to the probability measure $\nu(\cdot|X;\theta)$ on $\mathcal{U}$
4. The model generalized likelihood, $\mathcal{L} : 2^{\mathcal{Y}} \to [0, 1]$ is given
$$\mathcal{L}(A|X;\theta) := \nu(G^{-1}(A|X;\theta)|X;\theta)$$
5. $\mathcal{U}^*$ is the set of all predicted combinations of equilibria. That is $\mathcal{U}^* = \{G(\epsilon|X;\theta) : \epsilon \in \mathcal{U}\} \subset 2^{\mathcal{Y}}$
6. Optimal Transportation approach is considering a bipartite graph $\mathcal{G}(X;\theta)$ on $\mathcal{Y} \times \mathcal{U}^*$
   (a) Edges link $y \in \mathcal{Y}$ to $u \in \mathcal{U}^*$ if $y \in u$. So $(y, u) \in E(\mathcal{G}(X;\theta))$ if $y \in u$.
   (b) $Q(\cdot|X;\theta)$ is a probability distribution over $2^{\mathcal{Y}}$ with support $\mathcal{U}^*$ induced by the equilibrium correspondence.
   $$Q(u|X;\theta) = \nu(\{\epsilon : G(\epsilon|X;\theta) = u|X;\theta\})$$
   (c) Vertex $y \in \mathcal{Y}$ has weight $P(y|X)$ and each vertex $u \in \mathcal{U}^*$ has weight $Q(u|X;\theta)$

Theorem 1 shows that $\theta \in \Theta_I$ if and only if, for any subset A of $\mathcal{Y}$, we have $P(A|X) \leq Q(G^{-1}(A)|X;\theta)$, where $G^{-1}(A) = \{u \in \mathcal{U}^*|A \cap u \neq \theta\}$. Galichon and Genry show that it is equivalent to the existance of a joint probability Λ on $\mathcal{G}(\theta, X)$ with marginal distributions $P(\cdot|X)$ and $Q(\cdot|X;\theta)$.[4]

Theorem 3. *The parameter value θ belongs to the identified set iff there exists a probability on $\mathcal{Y} \times \mathcal{U}^*$ with support contained in $G(X;\theta)$ and with marginal probabilities $P(\cdot|X)$ and $Q(\cdot|X;\theta)$.*

One implication is easy to prove. Call U the random element with distribution Q (U is the "preimage" of $Q(u|X;\theta)$ from $(\Omega, \mathcal{F}, \mathbb{P}) \to 2^{\mathcal{Y}}$) If a joint probability Λ exists with all the required properties then

$$Y \in A \implies U \in G^{-1}(A)^5$$

so that $\mathbb{1}_{\{Y \in A\}} \leq \mathbb{1}_{\{U \in G^{-1}(A)\}}$, Λ-a.s. Taking expectations, gives $\mathbb{E}_\Lambda(\mathbb{1}_{\{Y \in A\}}) \leq \mathbb{E}_\Lambda(\mathbb{1}_{\{U \in G^{-1}(A)\}})$. Equivalently, $\mathbb{P}(A|X) \leq Q(G^{-1}(A)|X;\theta)$. The converse is more involved and relies on optimal transportation theory. A similar result is proved in Theorem 3 of Artstein (1983), based on an extension of the marriage lemma.

[4] I'm a bit lost on this part. I don't quite know what a probability is on a bipartite graph. I guess we just mean a joint probability distribution on the product space of the outcomes and the sets of outcomes. The theorem makes more sense intuitively looking at it from this view. For a consistent parameter, the rules relating the weights on one side with weights on the other are given above.

[5] How I think about this: Y is a map from ϵ onto $\mathcal{Y}$. U maps from ϵ to $\mathcal{U}^*$. Y gives the observed equilibria, U gives all the possible outcomes. A is a set of possible outcomes. The observed outcome must be a possible outcome. So if the observed outcome is in A, then t he set of possible outcomes must also intersect A. So $Y \in A \to U \in G^{-1}(A)$

Example (Family Bargaining Example cont.)**.** For the case of the family bargaining game

$$\mathcal{U}^* = \left\{ \{(0,0)\}, \{(0,1)\}, \{(1,0)\}, \{(1,1)\}, \{(0,1),(1,0)\} \right\}$$

The bipartite graph depicting this is in Figure 2(a), where $p_y, y \in \mathcal{Y}$ denotes $\mathbb{P}(y|X)$ and $q_u, y \in \mathcal{U}^*$ denotes $\nu(\{\epsilon : G(\epsilon|X;\theta) = u|X;\theta\})$. The existence of a joint probability on $\mathcal{Y} \times \mathcal{U}^*$ supported on $\mathcal{G}(X;\theta)$ with marginal probabilities $p_y, y \in \mathcal{Y}$ and $q_u, u \in \mathcal{U}^*$ can be represented graphically by a set of non-negative numbers attached to each edge of the graph that sum to 1 and such that the weight of each vertex is equal to the sum of the weights on the edges that reach it. For instance, a joint probability is denoted α_y^u for $(y,u) \in \mathcal{Y} \times \mathcal{U}^*$ and must satisfy $\alpha_y^u \geq 0$ for all $(y,u) \in \mathcal{Y} \times \mathcal{U}^*$, $\alpha_y^u = 0$ if $y \notin u$, $\sum_{y \in u} \alpha_y^u = 1$ and equalities such as $p_{01} = \alpha_{01}^{01} + \alpha_{01}^{01,10}$ and $q_{01,10} = \alpha_{01}^{01,10} + \alpha_{10}^{01,10}$.

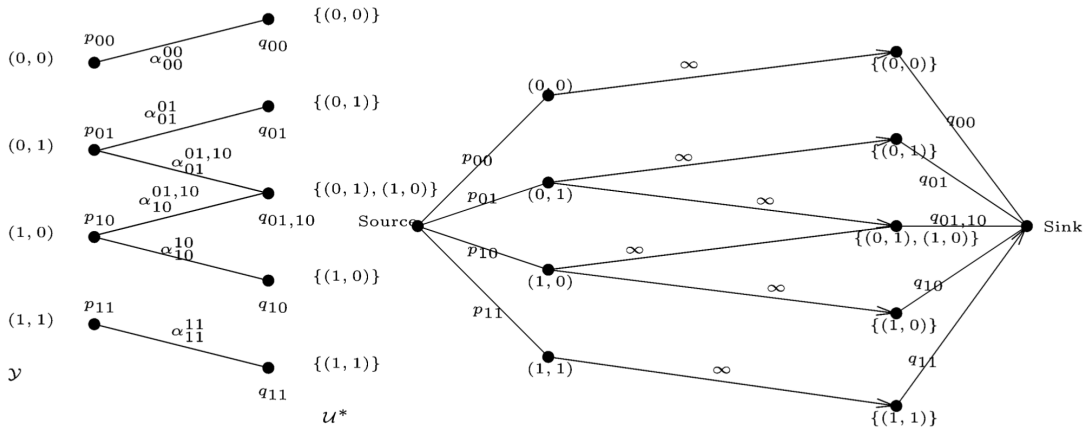


Figure 2: Family Game (continued), lifted from Galichon and Henry (2011)

Since the problem of computing the identified set has now been formulated as on involving the existence of a probability measure with given marginal distributions, one can appeal to efficient computational methods in the optimal transportation literature. The problem of sending $p_y, y \in \mathcal{Y}$ *units of a good* to $q_i, u \in \mathcal{U}^*$ *units in terminals*, $u \in \mathcal{U}^*$ at minimum cost of transportation, where costs are attached to each pair $(y,u) \in \mathcal{Y} \times \mathcal{U}^*$ is called an optimal transportation problem [6]. The problem at hand can be reduced to an optimal transportation problem where a pair (y,u) is assigned cost zero if it belongs to $\mathcal{G}(X;\theta)$ and one otherwise. There exists a joint law on $\mathcal{G}(x;\theta)$ with marginals P and Q (that is $\theta \in \Theta_I$) if and only if there is a zero-cost solution to the optimization problem thus defined. This minimum cost of transportation problem has an equivalent dual formulation as a maximum flow problem described in Figure 2(b). The edges in the graph with 0 cost in th minimum cost of transportation problem have infinite carrying capacity in the dual maximum flow problem. So, efficient maximum flow programs can be applied directly to the network described in Figure 2(b): mass flows from the source to the sink through the network.

[6] Basically, we have a total of 1 unit of "good" on the left hand ($\mathcal{Y}$) side, split over the elements of $\mathcal{Y}$ and 1 unit of "capacity" on the right hand side ($\mathcal{U}^*$) split over the units of $\mathcal{U}^*$. We need to get the goods on the left hand side to the right hand side, but can only use the paths $(y,u) \in \mathcal{G}(X;\theta)$). Because there is one unit of good on the LHS and 1 total unit of capacity on the RHS, all carrying units on the RHS are "filled". This is equivalent to there being a valid equilibrium selection mechanism.

3.3.3 Core-determining classes

Theorem 1 allows reduction of the problem of computing the identified set to that of checking a set of inequalities. However, the computational burden is only partially lifted, as the number of inequalities to check can be very large if the cardinality of the outcomes space is large. This section will analyze ways of reducing this remaining computational burden by eliminating redundant inequalities in the computation of the identified set. This is formalized with the concept of *core determining classes*, first introduced in Galichon and Henry (2006).

Definition 7 (Core Determining)**.** A class $\mathcal{A}$ of measurable subsets of $\mathcal{Y}$ is called *core determining* for the Choquet capacity $\mathcal{L}$ on $\mathcal{Y}$ if it sufficient to characterize the core of $\mathcal{L}$, i.e if a probability $Q \in \text{core}(\mathcal{L})$ when $Q(A) \leq \mathcal{L}(A)$ for all $A \in \mathcal{A}$.

A core-determining class of sets is sufficient to characterize the identified region Θ_I as summarized in the following proposition:

Proposition 1. *If $\mathcal{A}$ is core determining for the Choquet capacity $A \mapsto \mathcal{L}(A|X;\theta) = \nu(\epsilon : G(\epsilon|X;\theta) \cap A \neq \emptyset|X;\theta)$, X-a.s, and for all θ, then*

$$\Theta_I = \{\theta \in \Theta : \forall A \in \mathcal{A}, P(A|X) \leq \mathcal{L}(A|X), X\text{-}a.s\}$$

Challenge therefore becomes that of finding a core-determining class $\mathcal{A}$ in order to reduce the number of inequalities to be checked to the cardinality of $\mathcal{A}$. Return to the family bargaining example to illustrate:

Example (Family Bargaining cont.)**.** Return to the family bargaining game and consider some proposal for the computation of the identified set proposed in the literature. Call the *Singleton class* the class of singleton sets $\{\{(0,0)\},\{(0,1)\},\{(1,0)\},\{(1,1\}\}$. It is immediate to see that this class is not core determining in general. Indeed if ϵ has large enough support, the two equalities $\mathbb{P}(\{(0,1)\}) = \nu(G(\epsilon|\theta) \cap \{(0,1)\} \neq \emptyset|\theta) = \nu(\epsilon : \epsilon_1 \leq \theta, \epsilon_2 \geq -2\theta|\theta)$ and $P(\{(1,0\}) = \nu(G(\epsilon|\theta) \cap \{(1,0)\} \neq \emptyset|\theta)$ jointly imply $\mathbb{P}(\{(0,1),(1,0)\}) > \nu(G(\epsilon|\theta) \cap \{(0,1),(1,0)\} \neq \emptyset|\theta)$

Now show how to identify core-determining classes more generally to avoid case-by-case elimination of redundant inequalities. To this end, give general conditions under which one can find a core-determining class of low cardinality. Recall a subset A of an ordered set (equipped with $\succsim$) is said to be *connected* if any a such that $\sup A \succsim a \succsim \inf A$ belongs to A[7].

Assumption 2 (Monotonicity)**.** There exists an ordering $\succsim_{\mathcal{Y}}$ of the set of outcomes $\mathcal{Y}$ and an ordering $\succsim_{\mathcal{U}}$ of the set of latent variables $\mathcal{U}$ such that $G(\epsilon|X;\theta)$ is connected for all $\epsilon \in \mathcal{U}$, X-a.s, all θ, and $\sup G(\epsilon_2|X;\theta) \succsim_{\mathcal{Y}} \sup G(\epsilon_1|X;\theta)$ and $\inf G(\epsilon_2|X;\theta) \succsim_{\mathcal{Y}} \inf G(\epsilon_1|X;\theta)$ when $\epsilon_2 \succsim_{\mathcal{U}} \epsilon_1$. Both orderings are allowed to depend on the exogenous variables X,but the dependence is suppressed in the notation for clarity.[a]

[a] This assumption is related to monotone comparative statics in supermodular games

Example (Family bargaining cont.)**.** IN the family bargaining game, the orderings are very simple to construct. A lexicographic order on $\mathcal{Y}$ is suitable with $(0,0) \precsim_{\mathcal{Y}} (0,1) \precsim_{\mathcal{Y}} (1,0) \precsim_{\mathcal{Y}} (1,1)$. On $\mathcal{U}$ the ordering is related to predicted outcomes. All ϵ producing the same predicted outcomes will be equivalent, and the ordering on predicted outcomes is $\{(0,0)\} \precsim_{\mathcal{U}} \{(0,1)\} \precsim_{\mathcal{U}} \{(0,1),(1,0)\} \precsim_{\mathcal{U}} \{(1,0)\} \precsim_{\mathcal{U}} \{(0,1)\}$. Here $A_1 \precsim_{\mathcal{U}} A_2$ is a short-hand notation for $\epsilon_1 \precsim_{\mathcal{U}} \epsilon_2$ if $G(\epsilon_1|X;\theta) = A_1$ and $G(\epsilon_2|x;\theta) = A_2$.

For instance, taking ϵ_1 such that $G(\epsilon_1|X;\theta) = \{(0,1)\}$ and ϵ_2 such that $G(\epsilon_2|\theta) = \{(0,1),(1,0)\}$ we have $\sup G(\epsilon_1|\theta) = (0,1) \precsim_{\mathcal{Y}} (1,0) = \sup G(\epsilon_2|\theta)$ and $\inf G(\epsilon_1|\theta) = (0,1) = \inf G(\epsilon_2|\theta)$[8] This is illustrated in Figure 3.

[7] For example $[0,1]$ is connected while $(0,1)$ and $[0,1] \cup [2,3]$ are not

[8] Here it is useful to remember that the infima and suprema are being taken with respect to the orderings $\succsim_{\mathcal{U}}$ and $\succsim_{\mathcal{Y}}$.

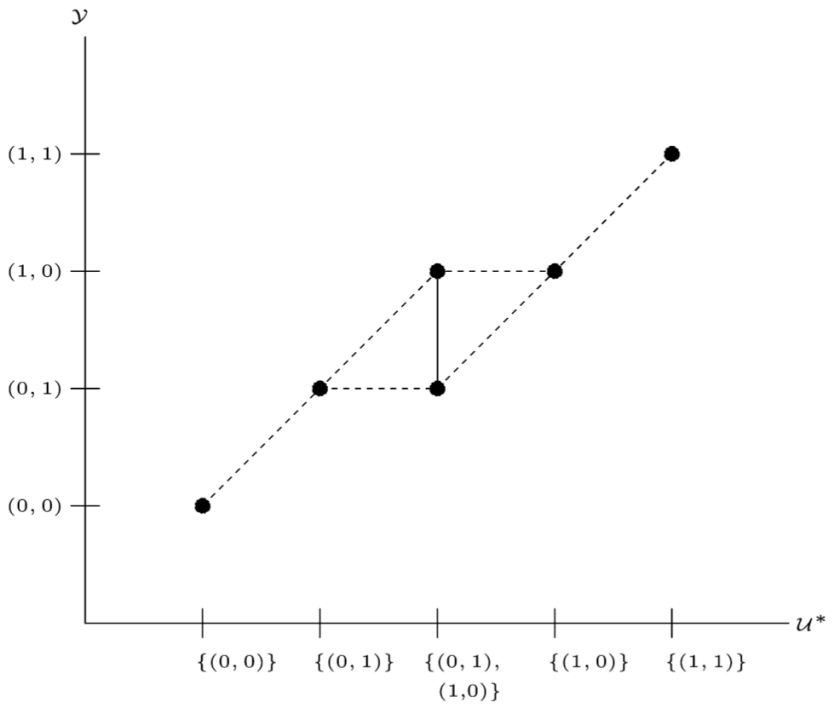


Figure 3: The monotonicity assumption in Assumption 2 is satisfied for this choice of orderings in the family bargaining example. The thick dots represent the correspondence $G(\cdot|\theta)$, $\mathcal{U}^*$ denotes the ordered set of combinations of equilibria.

This sets up the theorem, which is the main tool in the construction of core-determining classes, and hence in the compuation of the identified set under the monotonicity assumption.

Theorem 4. *Suppose Assumption 2 is satisfied with orderings $\succsim_{\mathcal{Y}}$ and $\succsim_{\mathcal{U}}$. Cal I the cardinality of $\mathcal{Y}$ and list outcomes (elements of $\mathcal{Y}$) in increasing order (w.r.t $\succsim_{\mathcal{Y}}$) as $y_1, \ldots, y_I$. Then $\mathcal{A} = (\{y_1, \ldots, y_i\}, \{y_i, \ldots, y_I\}, i = 1, \ldots, I)$ is core determining.*

Theorem 4 allows to reduce the cardinality of the power set $2^{\mathcal{Y}}$ to twice the cardinality of $\mathcal{Y}$ minus 2, since the inequality need not be checked on the whole set $\mathcal{Y}$.

3.4 Illustration: Oligopoly Entry with Two Types of Players

Now turn to a more substantive illustration of methods to compute the identified set. First, to show the operational usefulness of Theorem 4. To do so, consider the oligopoly entry game with two types of players presented in Appendix A of Berry and Tamer (2006).

The profit function of of Type 1 firms depends on the total number of firms in the market, but not on the type of those firms, while the profits of Type 2 firms depend both on the number and on the type of firms present in the market. The latent variable is the fixed cost f_1 for firms of Type 1 and f_2 for firms of type 2. Assume $(f_1, f_2) \sim \text{Unif}([0,1]^2)$. Model is simplified by assuming linearity of profits in the firm number as follows:

$$\Pi_1(Y_1, Y_2, X, f; \theta) = \alpha_0 + \alpha_1(Y_1 + Y_2) + \alpha_2 X - f_1$$
$$\Pi_2(Y_1, Y_2, X, f; \theta) = \beta_0 + \beta_1 + \beta_2 Y_2 + \beta_3 X - f_2$$

with $\alpha_1, \beta_1, \beta_2 < 0$ and $\beta_1 > \beta_1$. Set of observable outcomes is

4 A Geometric Approach to Inference in Set-Identified Entry Games; *Christian Bontemps, Rohit Kumar (JoE, 2020)*

Bontemps and Kumar (2020) is set to appear in Journal of Econometrics in 2020. They consider inference procedures for entry games with complete information. Complete the model with the unknown selection mechanism and characterize geometrically the set of predicted choice probabilities. A 2019 version of the paper can be found here.

4.1 Introduction

Paper provides an estimation procedure for empirical models of entry and market structure. Entry games are popular in the empirical Industrial Organization literatures 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. Popularized by the seminal works of Bresnahan and Reiss (1991a).

Econometrics 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 incomplete.

This paper completes the model with the selection mechanism $\eta(\cdot)$ and characterizes the set of predicted choice probabilities generated by the variation of $\eta(\cdot)$ in the space of admissible selection mechanisms. First contributions is to characterize more deeply the geometric structure of this set.

Set ends up being a convex polytope with many facets (because of focus on pure strategy equilibrium). This paper derived 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. Vertices are 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. 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. 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 of the the identified set - from 16 in a game with 3 players to more than 1 million in a game with 6 players.

4.2 Entry Game with N players

Formalize the entry game considered with N firms. First consider a model without explanatory variables.

4.2.1 Setup and Notations

Model Let N denote the number of firms that can enter any market. Following Berry (1992), 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 (action $a_{im} = 1$, doing so if their profit is positive. If $\pi_{im} = 0$, firms do not enter the market (take action $a_{im} = 0$). The profit function is assumed, without loss of generality, to be linear in the explanatory variables[1]

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

Following the literature, assume that $\alpha_i < 0$, i.e, the presence of more competitors decreases a firm's profit. Unobserved components $\epsilon_{im}, i = 1, \ldots, N$ are drawn from a known distribution (up to some parameter vector

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

γ). Econometrician does not observe their values but firms do.

For identification, we first need a scale normalization, and thus, assume that the variance of each shock ϵ_{im} is equal to 1. Denote the distribution of $\epsilon_m = (\epsilon_{1m}, \ldots, \epsilon_{Nm})^T$ and assume that the distribution is continuous with full support. Use notation θ for all the parameters in the model, and omit the subscript m for notational convenience. Assume $\theta \in \Theta \subseteq \mathbb{R}^l$.

Multiplicity of Pure Strategy eqm. For a given market, an outcome y is the vector of actions (in $\{0,1\}^N$) taken by the firms. There are 2^N possible outcomes. 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 firms playing action 1. There is 1 outcome with 0 active firms, N outcomes with 1 active firm, $d_k = \binom{N}{K}$ with K active firms, etc.

Globally, order the outcomes in $\mathcal{Y}$ first by their number of active firms, then according to the predefined order within each $\mathcal{Y}_K$:

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

Well known that the model has multiple equilibria, regions of realizations of ϵ in which one cannot uniquely predict each firms' action. Consequently, no one-to-one mapping between the collection of possible outcomes and the regions of ϵ given any parameter value θ. Consequently no one-to-one mapping between the collection of possible outcomes and the regions of ϵ given any parameter value θ.

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

Definition 1 (Equilibrium Selection Mechanism)**.** An equilibrium selection mechanism is a conditional probability $\eta(\cdot|\epsilon;\theta)$ such that the selected value of the outcome variable is actually an equilibrium predicted by the game. That is $\text{supp}(\eta(\cdot|\epsilon;\theta)) \subseteq G(\epsilon|\theta)$

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 is θ and selection mechanism is $\eta(\cdot)$. Partition this vector according to the partition of $\mathcal{Y}$ as

$$P(\theta,\eta) = \left(\underbrace{P_1^{(0)}(\theta,\eta)}_{P^{(0)}(\theta,\eta)}, \ldots, \underbrace{P_1^{(K)}(\theta,\eta), \ldots, P_{d_K}^{(K)}(\theta,\eta)}_{P^{(K)}(\theta,\eta)}, \ldots \underbrace{P_1^{(N)}(\theta,\eta)}_{P^{(N)}(\theta,\eta)} \right)^T \tag{2}$$

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

Another solution, 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 an unrestricted point, belongs to a convex set that is characterized. Different sets of values (θ,η) may generate the same point $P(\theta,\eta)$. Goal is to characterize which ones generate the true (read: observed) choice probability vector.

4.2.2 Choice Probabilities to Identified Set

Want to characterize the set of predicted choice probabilities. To do so, need to understand the multiplicity structure and characterize it. Then derive a parameterization of the set.

Regions of Multiple Equilibria Specification ensures that multiple equilibria only involve outcomes with the same number of active firms, i.e within $\mathcal{Y}_K$. Therefore, focus on subsets of outcomes $S \subseteq \mathcal{Y}_K$ to characterize the multiple equilibria regions. 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 $\mathcal{R}_S^{(K)}(\theta)$.[2] $\mathcal{R}_S^{(K)}$ is called a multiple equilibria region. Denote by $S^{(K)}$ the collection of subsets S of $\mathcal{Y}_K$ in multiplicity[3].

$$S^{(K)} = \left\{ S \subset \mathcal{Y}_K : |S| \geq 2 \text{ and } S \text{ is in multiplicity} \right\}$$

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

Now present necessary and sufficient condition for S to be in multiplicity.

$$N_0(S) = \{\text{Set of indices of firms that always play action 0 across } S\}$$
$$N_1(S) = \{\text{Set of indices of firms that always play action 1 across } S\}$$

Further define $n_0(S) = |N_0(S)|$ and $n_1(S) = |N_1(S)|$, the cardinalities of the sets above. For now, suppress the dependence on S. With N_0 and N_1 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 $K - n_1$ which play action $a_{im} = 1$ among the $N - n_0 - n_1$ remaining ones. S should contain all these possibilities to be in multiple equilibria.

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

For the 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, $\binom{4-1-1}{2-1} = 1$. Proof of proposition 1 also characterizes the region $\mathcal{R}_S^{(K)}(\theta)$. Proposition 1 can be used to count the number of multiple equilibria regions:

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*

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

[a] We can think of this as, for a given K firms that are active in the quilibrium, choose n_1 to "always" be in and n_0 to "always" be out. So long as $n_0 + n_1 < N$ and $n_1 < K$, there is garunteed to be a region of mulitple equilibria corresponding to this by Proposition 1.

With $K = 1$, the number of regions of multiple equilibria is $\sum_{n=0}^{N-2} \binom{N}{n}$, all possible combinations of more than two outcomes. However, Table 1 shows that the number of regions for the various values of N and K is generally far less than all the possible combinations. So the parameterization of the set of predicted choice probabilities is of a much lower dimension than one would have expected.

The set of predicted choice probabilities Also define the subset of $S^{(K)}$ that contains one specific outcome $y_j^{(K)}$ as

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

[2] In the Galichon Henry (2011) notation, this means that $S = G(\epsilon|\theta)$ for some $\epsilon \in \mathcal{U}$. Then $\mathcal{R}_S^{(K)}(\theta) = \{\epsilon \in \mathcal{U} : G(\epsilon|\theta) = S\}$

[3] The maximum number of such subsets is equal to $2^{d_K} - d_k - 1$

[4] This just means that, there exists a value of $(\theta, \epsilon) \in \Theta \times \mathbb{R}^N$ such that both outcomes in S_2 are simultaneously equilibria of the game. That is, given θ and ϵ, both $(1,1,0,0)^T$ and $(1,0,1,0)^T$ are equilibria. However, there is no value of (θ, ϵ) such that both $(1,1,0,0)^T$ and $(0,0,1,1)^T$ are both simultaneously equilibrium outcomes.

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	3	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	14	198	32761
	5	6	57	57

Table 1: Counting the number of multiple equilibria regions [Lifted from paper]

Following Berry and Tamer (2007) and Galichon and Henry (2011), can calculate the probability of observing outcome $y_j^{(K)}$. This probability depends on the parameter vector θ and the selection mechanism η. Specifically, letting $U_j^{(K)}(\theta)$ be the region in the support of ϵ which uniquely predicts the outcome $y_j^{(K)}$:

$$P_j^{(K)}(\theta,\eta) = \int_{U_j^{(K)}(\theta)} dF(\epsilon;\theta) + \sum_{S\in S_j^{(K)}} \int_{R_S^{(K)}(\theta)} \eta\left(y_j^{(K)}|\epsilon;\theta\right) dF(\epsilon;\theta) \tag{3}$$

Denote by

$$\Delta_j^{(K)}(\theta) := \int_{U_j^{(K)}(\theta)} dF(\epsilon;\theta) \;\text{ and }\; \Delta_S^{(K)}(\theta) := \int_{R_S^{(K)}(\theta)} dF(\epsilon;\theta)$$

Let $A(\theta)$ be the set of $P(\theta;\eta)$ generated by the variation of η in $\mathcal{E}$[5] and let B_K^θ be the set of $P^{(K)}(\theta,\eta)$ generated by the variation of η in $\mathcal{E}$, for $K=0,\ldots,N$. Formally

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

Equation 3 can then be viewed as a parameterization of the sets $A(\theta)$ and $B_K(\theta)$ where the "parameters" are the regions $\mathcal{R}_S^{(K)}(\theta)$[6].

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 and selection mechanism. The identified set Θ_I is defined as the collection of points θ such that P_0 can be rationalized with a selection mechanism

$$\Theta_I := \left\{\theta\in\Theta : \exists\eta\in\mathcal{E}, P_0=P(\theta,\eta)\right\} \tag{4}$$

The following is easily verified and intuitive

$$\theta\in\Theta_I \iff P_0\in A(\theta) \tag{5}$$

So, can study Θ_I by studying the structure of $A(\theta)$. The following result holds:

[5]Recall $\mathcal{E}$ is the set of all valid equilibrium selection mechanisms.

[6]This means that elements of the sets $A(\theta)$, $B_K(\theta)$ are characterized/fully determined by their corresponding regions $\mathcal{R}$.

Proposition 3. *$A(\theta)$ is a convex subset of $\mathbb{R}^{2^N}$, each $B_K(\theta)$ is a convex subset of $\mathbb{R}^{d_K}$, and*

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

The convexity of $A(\theta)$ is a general feature of an entry game and does not depend on this specification. The specific structure, the direct product nature, comes from the specification in Equation (1). Structure simplifies some of the results to follow.

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 in the cube, Cub_K, defined by

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

This follows simply from the breakdown of $P_j^{(K)}$ in equation (3), the following definitions of $\Delta_j^{(K)}$ and $\Delta_S^{(K)}(\theta)$, and that $0 \leq \eta(\cdot) \leq 1$.

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, can decompose this condition to

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

4.2.3 The Support Function and a First Selection of Moment Inequalities

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

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

for all directions $q \in \mathbb{R}^d$. This is depicted visually in Figure 1, below. Generally, the domain of $\delta(\cdot)$ is restricted to $\mathbb{S}^{n-1}$ The support function of a convex set in a given direction locates the supporting hyperplane

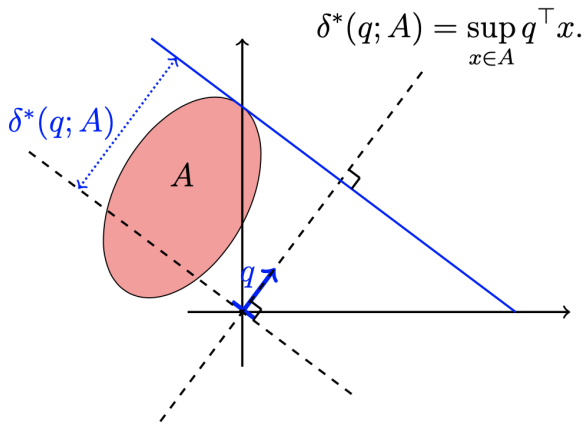


Figure 1: The support function [Lifted from Bontemps and Kumar (2020)]

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, there are only a finite number of (linear) inequalities needed 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 Rockafeller (1970) and Proposition 3, the identified set is characterized by the following inequalities.

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

Breaking this down. P_0 is the observed "true" vector of outcome probabilities. $A(\theta) \subset \mathbb{R}^{2^N}$ is the set of probabilities that can be rationalized by an equilibrium selection mechanism in the model with parameter vector θ. $P_0^{(K)} \in \mathbb{R}^{d_K}$ is the sub-vector of probabilities for outcomes with K active firms, where $d_K = \binom{N}{K}$, the number of possible outcomes with K active firms. $B_K(\theta) \subset \mathbb{R}^{d_K}$ is the set of sub-vectors of observed probabilities for outcomes with K active firms that can be rationalized by an equilibrium selection mechanism in a model generated by parameter vector θ.

Now turn to the calculation of the support function of $B_K(\theta)$ for any K. Make the following notations:

1. Let $q_K \in \mathbb{R}^{d_K}$ be a given direction. Assume the following order among the coordinates of q_K: $q_{i_1,K} \geq q_{i_2,K} \geq \ldots, \geq q_{i_{d_K},K}$.
2. Partition $S^{(K)}$, the collection of subsets of outcomes with K active firms in multiplicity, as follows:
   (a) Denote $\mathcal{O}_{i_1}^{(K)} = S_{i_1}^{(K)}$[7], the elements of $S^{(K)}$ which contain the outcome $y_{i_1}^{(K)}$[8].
   (b) Denote by $\mathcal{O}_{i_2}^{(K)} \subset S_{i_2}^{(K)}$ the subset of elements of $S_{i_2}^{(K)}$ that are not in $\mathcal{O}_{i_1}^{(K)}$, i.e $\mathcal{O}_{i_2}^{(K)} = S_{i_2}^{(K)} \backslash S_{i_1}^{(K)}$.
   (c) Continue on in this fashion for each $i_j, j \in 3, \ldots, d_K$ defining

$$\mathcal{O}_{i_j}^{(K)} = S_{i_j}^{(K)} \setminus \bigcup_{k<j} S_{i_k}^{(K)}$$

Note that the construction of the outcomes $\mathcal{O}_j^{(K)}$ is lined to the order of the components of q_k. Now provide a closed form expression for the support function in this direction (Prop 4, below).

[7] Remember $S_j^{(K)} = \{S \in S^{(K)} : y_j \in S\}$

[8] $y_{i_1}^{(K)}$ being the outcome with K active firms that has the highest probability assigned to it by $q_K \in \mathbb{R}^{d_K}$

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

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

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 if indices is any permutation of the vector of indices $(1, 2, \ldots, d_K)$. As such $B_K(\theta)$ has at most $d_K! = \binom{N}{K}!$ vertices.

Each extreme point of $B_K(\theta)$ 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 refions, and we saw in Proposition 2 that this number is much smaller than $2^{\binom{N}{K}} - \binom{N}{K} - 1$. Consequently, the parameterization of $B_K(\theta)$ is numerically tractable for moderate values of N. Furthermore, each non-zero value $\Delta_S^{(K)}(\theta)$ can be easily calculated or simulated from the knowledge of the distribution of ϵ.

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

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

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

Results come from the specific characterization of $A(\theta)$ in proposition 3. The last proposition, combined with equation (7), is the basis of the inference problem. 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. Now explicit this set of directions, first for the $B_K(\theta)$'s and then 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 $2^{d_K} - 1$ elements in $\mathcal{Q}_K$. The next proposition shows that is sufficient to check the inequalities in $\mathcal{Q}_K$, fo all K, to characterize the identified set:

Proposition 6. *Let $P_0 \in \mathbb{R}^{2^N}$ denote the observed vector of probabilities. Then*

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

Remark Already mentioned that the 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 a 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, 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.

Optimal Transport, Random Sets, or Completion of the Model Approach consists in characterizing the set $A(\theta)$ through its support function and 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. 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 optimal transportation theory and the notion of core determining classes 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 the 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 $N \geq 6$; have more than 1 million inequalities to check. 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, but 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 methods are intended to derive a sufficient set of inequalities satisfied by the parameters in a specific manner. Each method has its specificities. However, this approach allows us to go deeper into the geometric analysis of the set $A(\theta)$ and this is the objective of the next section.

4.3 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. In this section, 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). Second consists of exploiting the geometry to propose a geometric selection procedure of the inequalities without having to evaluate all of them.

4.3.1 Deriving a Core Determining Class of an Entry Game

Core determining classes yields a collection of non-redundant moment inequalities that are sufficient to sharply characterize the identified set Θ_I. Provide a characterization of the core determining class in an entry game from the geometric study of the multiplicity structure of the model.

As a reminder,

Definition 2 (Choquet Capacity)**.** $\mathcal{L}: 2^{\mathcal{Y}} \to \mathbb{R}$ is a *Choquet Capacity* if it is
1. *normalized*: $\mathcal{L}(\emptyset)=0$ and $\mathcal{L}(\mathcal{Y})=1$, and
2. *monotone*: $\mathcal{L}(C) \leq \mathcal{L}(B)$, for any $C \subseteq B \subseteq \mathcal{Y}$

Definition 3 (Core Determining)**.** $\Omega \subset 2^{\mathcal{Y}}$ is called core determining for the Choquet Capacity $\mathcal{L}$ on $\mathcal{Y}$ if, for an arbitrary random variable X taking values on $\mathcal{Y}$ and associated law $\mathbb{P}$ (arbitrary probability distribution $\mathbb{P}$ on $\mathcal{Y}$):

$$\mathbb{P}(C) \leq \mathcal{L}(C), \forall C \in \Omega \implies \mathbb{P}(C) \leq \mathcal{L}(C), \forall C \in 2^{\mathcal{Y}}$$

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

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

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

quantities that are positive, and $\mathcal{L}(\mathcal{Y}) = 1$.

Define the concept of connectedness, which is useful for the exposition, introduced by Galichon and Henry (2011). For a subset $C_K \subset \mathcal{Y}_K$, define the (undirected) graph generated by C_K as $\Gamma_{C_K} = (C_K, E)$. 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.

Definition 4 (Well Connectedness)**.** A subset $C_K \subseteq \mathcal{Y}_K$ is called well connected in $\mathcal{Y}_K$ if $\mathcal{Y}_K \setminus C_K$ is connected in the graph $\Gamma_{\mathcal{Y}_K \setminus C_K}$ [a].

[a] This is the subgraph obtained by deleting all nodes in C_k and all associated edges.

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 complement of C_K.

The graph $\Gamma_{\mathcal{Y}}$ is not connected, as there is no multiplicity between $\mathcal{Y}_K$ and $\mathcal{Y}_{K'}$ for $K \neq K'$. The $\Gamma_{\mathcal{Y}_K}$ is a component of $\Gamma_{\mathcal{Y}}$[1] Collect all well-connected subsets of $\mathcal{Y}_K$ as

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

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 submodular 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 IV model. Next proposition provides a complete characterization of a core determining class for entry model through a necessary and sufficient condition.

Proposition 7. *A collection Ω of subsets of $\mathcal{Y}$ ($\Omega \subseteq 2^{\mathcal{Y}}$) is core determining for $\mathcal{L}$ as defined in equation* (10) *if and only if $\Omega = \{\Omega_K : K = 0, 1, \ldots, N\}$.*

This selection, however, does not significantly reduce the number of non-redundant moment inequalities in the entry game. For example, when $N = 6$ and $K = 3$, it eliminates fewer than $30,000$ inequalities from a total of $2^{20} - 1 = 1,048,575$.

4.3.2 A Geometric Selection Procedure

The core determining class is a useful concept since it eliminates redundant inequalities. However, it does not significantly reduce the number of inequalities in the entry game. This section presents a geometric selection procedure that fully exploits the geometery of the sets $B_K(\theta)$. Procedure first selects the extreme point of the set that seems closes to the vector $P_0^{(K)}$ and then only evaluates only the ineqaulities associated with this extreme point (i.e tests the directions that are the outer normal vector of the supporting hyperplans of $B_K(\theta)$ at this point, for each $K = 0, \ldots, N$)[2].

Following Proposition 4, an extreme point is determined by an order in the coordinates (note that, a priori, two different orders couldlead to the same physical point). The first part of the algortihm is intended to determine this order in a recursive manner by exploting the position of $P_0^{(K)}$ with respect to the cube Cub_K which contains $B_K(\theta)$.

Explain the steps in non-technical detail here and then formalize in the appendix.

Local Moment Selection: The 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. Them, determine which coordinate of $P_0^{(K)}$ is the furthest fromt he center of the cube.

[1] Components have no edges between them.

[2] By extreme point I believe the authors mean a point $x \in A(\theta)$ at which the support function $\delta(P_0, A(\theta)) = x^T P_0$

2. Assume this is the j^{th} coordinate.
   (a) If it is on the maximum side, the extreme point is of the type $E^{(K)}_{j,?,\ldots,?}(\theta)$ and now have to determine the next component. To do so, project $P_0^{(K)}$ on the face and 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, know that extreme point will be of the form $E^{(K)}_{?,\ldots,?,j}(\theta)$. Now have to determine the next component. To do so, project $P_0^{(K)}$ on this face and repeat the previous calculation, taking into accound 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 $E^{(K)}_{i_1,i_2,\ldots,i_{d_K}}$ is determined, can focus on the directions of the local supporting hyperplans. Let the d_K directions $e_{i_1}, e_{i_1,i_2}, \ldots, e_{i_1,i_2,\ldots,i_{d_K}}$, where the components are equal to 1 and indices are subscrips 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 doens't provide a sharp characterization of $B_K(\theta)$ unless $K = 1$ or $K = N-1$, but, however, provides an important refinement with respect to the existing method of Ciliberto and Tamer (2009).

Procedure selects which moments among the $2^{d_K} - 1$ are potentially binding without having to evaluate all of them. Selection is based on the spatial location of the point $P^{(K)}$ and exploits the geometery of the set $B_K(\theta)$

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

4.4 Estimation and Inference

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

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

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

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

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. The set $\mathcal{G}$ collects all the directions needed to sharply identify the identified set.

Test based on $T_\infty(\cdot)$ is infeasible because P_0 is not observed. Now characterize the feasible test statistic and its asymptotic distribution. Throughout this section, assume observe a sample of M i.i.d markets in which the same N firms compete.

4.4.1 The Asymptotic Distribution of the Test Statistics

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

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

Want asymptotic results to be valud, not only for the true probability, but also uniformly in the neighborhood around the true probability. Impose the following uniform integrability condition:

Assumption 1. The class $\mathcal{P}$ of probability distributions satisfies

$$\lim_{\lambda\to\infty}\sup_{P\in\mathcal{P}}\sup_{j\in\{1,\dots,2^N\}}\mathbb{E}_P\left[\left(\frac{\mathbb{1}(Y=y_j)-\mu_j(P)}{\sqrt{\mu_j(P)(1-\mu_j(P))}}\right)^2\mathbb{1}\left\{\left|\frac{\mathbb{1}(Y=y_j)-\mu_j(P)}{\sqrt{\mu_j(P)(1-\mu_j(P))}}\right|>\lambda\right\}\right]=0 \tag{11}$$

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

Assumption 1 ensures the uniform convergence of test statistic over the class of probability distributions. This condition is satisified over the class of probability distributions such that $\mu_j(P)\geq\epsilon$ for each j and some $\epsilon>0$.

NOTE: Rest of this discussion seems to follow standard asymptotic theory with the number of inequalities being tested remaining fixed as $n\to\infty$. This seems like the wrong setting.

Interested in covering a specific point in the identified set with some pre specified probability

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

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

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

A point θ belongs to the confidence interval if the test based on $\xi_M(\theta)$ is not rejected. 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*

$$\begin{cases}\xi_M(\theta)\rightsquigarrow\min q\in Q_\theta\frac{q^TZ}{\sqrt{q^T\Sigma q}} & \text{if } P_0\in A(\theta)\\ \xi_M(\theta)\to_{a.s}-\infty & \text{if } P_0\notin A(\theta)\end{cases}$$

Under Assumption 1, these results are uniformly valid over $P\in\mathcal{P}$.

Asymptotic distribution only depends on θ through Q_θ

5 Action-Graph Games *Albert Xin Jiang, Kevin Leyton-Brown, Navin A.R Bhat (GEB, 2011)*

Jiang et al. (2011) appeared in *Games and Economic Behavior* in 2011. It can be found online here.

5.1 Introduction

Simultaneous action games have received considerable study, which is reasonable as these games are fundamental. Most of the game theory literature presumes that simultaneous action games will be represented in normal form. This is problematic because, in many domains of interest, the number of players and/or the number of actions per player is large. In normal form representation, the game's payoff function is stored as a matrix with one entry for each player's payoff under each combination of all players' actions. As a result, the size of the representation grows exponentially with the number of players.

Fortunately, most games of practical interest have highly-structured payoff functions, and thus, it is possible to represent them compactly. Intuitively, this helps to explain why people are able to reason about these games in the first place: understand the payoffs in terms of simple relationships rather than in terms of large lookup tables. One thread of recent work has explored game representations that are able to succinctly describe games of interest. For example, the extensive form allows games with temporal structure to be encoded in exponentially less space than the normal form. In what follows, however, concentrate on game representations that are compact even for simultaneous-move games of perfect information.

Perhaps the most influential class of compact game representations is that which exploits strict independences between players' utility functions. Class includes graphical games, multi-agent influence diagrams, and game nets. Focus on the first of these.

Consider a graph in which nodes correspond to agents and an edge from one node to another represents the proposition that the first agent is able to affect the second agents payoff's. If every node in the graph has a small in-degree, the graphical game has a compact representation (exponentially smaller than it's induced normal form). Of course, there are many ways of representing games compactly. What makes graphical games important is the fact that computational questions about these games can be answered by algorithms whose running time depends on the size of the representation rather than the size of the induced normal form. To state one fundamental property, it is possible to compute an agent's expected utility under an arbitrary mixed strategy profile in time polynomial in the size of the graphical game representation.

This property implies that a variety of algorithms for computing game-theoretic quantities of interest, such as sample Nash and correlated equilibrium, can be made exponentially faster for graphical games without introducing any change in the algorithm's behavior or output. Property implies that a variety of algorithms for computing game-theoretic quantities of interest, such as sample Nash and correlated equilibrium can be made exponentially faster for graphical games.

A drawback of the graphical games representation is that it only helps when there exist agents who *never* affect some other agent' utilities. Unfortunately, many games of interest lack any structure of this kind. For example, nontrivial symmetric games are cliques when represented as graphical games. Another useful form of structure not generally captured by graphical games is dubbed *anonymity*; it holds when agents utility only depends on the number of agents who took each action, rather than

References

Beresteanu, A., Molchanov, I., and Molinari, F. (2011). Sharp identification regions in models with convex moment predictions. *Econometrica*, 79(6):1785–1821.

Beresteanu, A. and Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814.

Berry, S. and Tamer, E. (2007). *Identification in Models of Oligopoly Entry*, volume 2 of *Econometric Society Monographs*, page 46–85. Cambridge University Press.

Bontemps, C. and Kumar, R. (2020). A geometric approach to inference in set-identified entry games. *Journal of Econometrics*.

Bontemps, C., Magnac, T., and Maurin, E. (2012). Set identified linear models. *Econometrica*, 80(3):1129–1155.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2018). Inference on Causal and Structural Parameters using Many Moment Inequalities. *The Review of Economic Studies*, 86(5):1867–1900.

Chesher, A. and Rosen, A. M. (2017). Generalized instrumental variable models. *Econometrica*, 85(3):959–989.

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

Galichon, A. and Henry, M. (2006). Inference in incomplete models.

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

Jiang, A. X., Leyton-Brown, K., and Bhat, N. A. (2011). Action-graph games. *Games and Economic Behavior*, 71(1):141 – 173. Special Issue In Honor of John Nash.

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