1 Entry Models with Multiple Equilibria

In this set of lecture notes we introduce the ideas related to partial identification in structural econometric models. As a motivating example, we will reconsider the static two-firm entry game. Despite its simplicity, it is strategically nontrivial because the entry choices of competing firms are interdependent (ie. entry choice of firm 1 affects profits of firm 2).

As we will see, in this simple entry game, multiple equilibria are a typical problem. A literature has pointed out how, typically, the possibility of multiple equilibria in the underlying game leads to the *partial identification* of the structural model parameters.[1] This means that there are multiple values of the structural model parameters which are consistent with the observed data. Econometrically, the estimating equations in these types of settings typically take the form of "moment inequalities", and a very large literature has developed regarding inference with moment inequalities. These lecture notes will cover these topics.

Throughout, we will employ a two-firm entry game as the running example. First, we focus on games where the moment (in)equalities are generated by "structural" errors (ie. those observed by the firms, but not by the econometrician). A typical paper here is Ciliberto and Tamer's (2009) analysis of entry in airline markets. Second, we consider the case where the moment (in)equalities are generated by non-structural, expectational errors, which are not known by agents at the time that their decisions are made. This follows the approach taken in Pakes, Porter, Ho, and Ishii (2007).

1.1 Entry games with structural errors

Consider a simple 2-firm entry model. Let $a_i \in \{0, 1\}$ denote the action of player $i = 1, 2$. The profits are given by:

$$\Pi_i(s) = \begin{cases} \beta' s - \delta a_{-i} + \epsilon_i, & \text{if } a_i = 1 \\ 0 & \text{otherwise} \end{cases}$$

s denotes market-level control variables. Entry choices are interdependent, in the sense that, firm 1's profits from entering (and, hence, decision to enter) depend on whether firm 2 is in the market.

As before, the error terms ϵ_i are assumed to be observed by both firms, but not by the

[1] See, for instance, Tamer (2003), Ciliberto and Tamer (2009), Beresteanu, Molchanov, and Molinari (2011), Galichon and Henry (forthcoming).

econometrician. This is a "perfect information" game. We also consider "incomplete information" games below.

For fixed values of the errors $\epsilon \equiv (\epsilon_1, \epsilon_2)$ and parameters $\theta \equiv (\alpha_1, \alpha_2, \beta_1, \beta_2)$, the Nash equilibrium values a_1^*, a_2^* must satisfy best-response conditions. For fixed (θ, ϵ), the best-response conditions are:

$$\begin{aligned} a_1^* = 1 \Leftrightarrow &\quad \Pi_1(a_2^*) \geq 0 \\ a_1^* = 0 \Leftrightarrow &\quad \Pi_1(a_2^*) < 0 \\ a_2^* = 1 \Leftrightarrow &\quad \Pi_2(a_1^*) \geq 0 \\ a_2^* = 0 \Leftrightarrow &\quad \Pi_2(a_1^*) < 0. \end{aligned}$$

For some values of parameters, there may be multiple equilibria.

[INSERT PICTURE HERE]

Given this setup, we derive the following inequalities for the probabilities of the four entry outcomes:

- $P_{10}^U(\beta, \delta) \equiv [1 - \Phi(-\beta' s)][\Phi(\delta - \beta' s)] \geq Pr[(1,0)|s] \geq [1 - \Phi(-\beta' s)]\Phi(-\beta' s) + [1 - \Phi(\delta - \beta' s)][\Phi(\delta - \beta' s) - \Phi(-\beta' s)] \equiv P_{10}^L(\beta, \delta)$
- $P_{01}^U(\beta, \delta) \equiv [1 - \Phi(-\beta' s)][\Phi(\delta - \beta' s)] \geq Pr[(0,1)|s] \geq [1 - \Phi(-\beta' s)]\Phi(-\beta' s) + [1 - \Phi(\delta - \beta' s)][\Phi(\delta - \beta' s) - \Phi(-\beta' s)] \equiv P_{01}^L(\beta, \delta)$
- $[\Phi(-\beta' s)]^2 = Pr[(0,0)|s]$
- $[1 - \Phi(\delta - \beta' s)]^2 = Pr[(1,1)|s]$

Agnosticism, multiple equilibrium, and partial identification: a thought experiment. Why does multiple equilibria go hand-in-hand with partial identification? Consider a thought experiment, where John and Jill are given the same dataset, on entry outcomes from a two-firm entry game, played across a large number of identical markets. John and Jill agree on the model, but disagree about the equilibrium selection procedure. Let's say John believes that, when there are multiple equilibria, the $(0,1)$ outcome always obtains, so that $Pr[(0,1)|s] = P_{01}^U$ and $Pr[(1,0)|s] = P_{10}^L$. Jill believes, however, that in the multiple equilibria region, the two firms flip a coin so that $(0,1)$ and $(1,0)$ occur with 50-50 odds, so that $Pr[(0,1)|s] = 0.5(P_{01}^L + P_{01}^U)$ and $Pr[(1,0)|s] = 0.5(P_{10}^L + P_{10}^U)$.

Now they take the data and estimate the model parameters under their assumptions. (For instance, they could run maximum likelihood.) Obviously, John and Jill will obtain different estimates of the parameters (β, δ). Who is right? As an agnostic observer, you must conclude that *both* are right. Therefore, in this multiple equilibria setting, agnosticism about the equilibrium selection rule drives the partial identification of the model parameters.

1.2 Deriving moment inequalities

Define the mutually exclusive outcome indicators:

$$\begin{aligned} Y_1 &= \mathbf{1}(a_1 = 1, a_2 = 0) \\ Y_2 &= \mathbf{1}(a_1 = 0, a_2 = 1) \\ Y_3 &= \mathbf{1}(a_1 = 0, a_2 = 0) \\ Y_4 &= \mathbf{1}(a_1 = 1, a_2 = 1). \end{aligned} \tag{1}$$

We observe a dataset $\vec{Y_t} = \{Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t}\}$ for a series of markets $t = 1, \ldots, T$. From this data, we can estimate the outcome probabilities $\hat{P}_{00}, \hat{P}_{01}, \hat{P}_{10}, \hat{P}_{11}$ (we ignore market-specific covariates s for now). These estimates can be plugged into the probability inequalities above, leading to *moment inequalities* which define the identified set of parameters (β, δ).

The identified set as defined by these moment inequalities, is not "sharp": don't impose joint restrictions in inequalities. Specifically, if you are at upper bound of the first equation, you cannot be at upper bound of second equation. That is, you need to impose an additional equation on

$$P[(0,1) \cup (1,0)|s] = P^U_{10}(\beta, \delta) + P^L_{01}(\beta, \delta) = P^U_{01}(\beta, \delta) + P^L_{10}(\beta, \delta).$$

There are alternative ways around this multiple equilibrium problem. Instead of modeling events $Y_1 = 1$ and $Y_2 = 1$ separately, we model the aggregate event $Y_5 \equiv Y_1 + Y_2 = 1$, which is event that *only one firm* enters. In other words, just model likelihood of *number of entrants* but not identities of entrants. Indeed, this was done in Berry's (1994) paper.

1.2.1 Ciliberto and Tamer (2009) paper

- Describe entry patterns in airline markets.
- Markets $m = 1, \ldots, n$: each market is a *city-pair*, with characteristics X_m.
- Potential entrants in each market $k = 1, \ldots, K$.

- Outcomes of interest: $\vec{y}_m = \{y_1^m, \ldots, y_K^m\}$ with $y_k^m = \mathbf{1}(\text{firm } k \text{ enters market } m)$. Let the vector $\mathbf{y}$ index all the possible outcome configurations for $\vec{y}$. For instance, with $K = 3$, then $\mathbf{y}$ indexes the events $\{(111), (110), (101), (011), (100), (010), (001), (000)\}$.

- For each covariate vector X (includes both market and firm covariates), Nash equilibrium behavior in entry game yields the moment inequalities

$$H_1(\theta, X) \leq P(\mathbf{y}|X) \leq H_2(\theta, X).$$

- Define identified set:

$$\Theta_I = \{\theta : \text{inequalities above are satisfied for all } X\}.$$

- Work with objective function

$$Q(\theta) = \int [||(P(\mathbf{y}|X) - H_1(\theta, X))_-|| + ||(P(\mathbf{y}|X) - H_2(\theta, X))_+||dF_X]$$

  with sample analogue

$$Q_n(\theta) = \frac{1}{n}\sum_{m=1}^{n}\left[||(\hat{P}(\mathbf{y}|X_m) - H_1(\theta, X_m))_-|| + ||(\hat{P}(\mathbf{y}|X_m) - H_2(\theta, X_m))_+||\right]$$

  where $\hat{P}(\mathbf{y}|\cdot)$ denotes estimates of the outcome probabilities, as functions of X.

- $P(\mathbf{y}|X_m)$ should be estimated nonparametrically, or flexibly parametrically. (Akin to choice probability estimation in dynamic models.)

- Typically, unless number of firms K is very small, the upper and lower bound probabilities $H_1(\theta, \cdot)$ and $H_2(\theta, \cdot)$ do not have a convenient closed form, and will need to be simulated, for different values of θ.

- Discuss inference below.

- Results

1.3 Entry games with expectational errors

In contrast to the above, Pakes, Porter, Ho, and Ishii (PPHI) derive the moment inequalities directly from the optimality conditions. By allowing more general error terms, a large variety of moment inequalities can be generated. We illustrate this approach again for the entry example.

Nash equilibrium In the two-firm entry game, if the actions (a_1^*, a_2^*) are observed, then the inequalities for a Nash equilibrium are

$$\begin{aligned} E\left[\pi_1(a_1^*, a_2^*, z)|\Omega_1\right] - E\left[\pi(a, a_2^*, z)|\Omega_1\right] > 0, \text{ for } a \neq a_1^* \\ E\left[\pi_2(a_1^*, a_2^*, z)|\Omega_2\right] - E\left[\pi(a_1^*, a, z)|\Omega_2\right] > 0, \text{ for } a \neq a_2^* \end{aligned} \tag{2}$$

These conditions are from the optimizing firms' point of view, so in order for the expectations above to be nontrivial, implicitly there are some variables in z, which are not observed by the firms (ie. not in the information sets Ω_1, or Ω_2).

Accordingly, PPHI parameterize (for all i, a_1, and a_2)

$$\pi_i(a_1, a_2, z) = r_i(a_1, a_2, z; \theta).$$

In the above, $r_i(a_1, a_2, z; \theta)$. is a particular functional form for firm i's counterfactual profits under action profile (a_1, a_2), which is assumed to be known by researchers, up to the unknown parameters θ.[2]

Hence, plugging into the equilibrium inequalities, we have the conditional moment inequalities which we can use to estimate θ:

$$\begin{aligned} E\left[r_1(a_1^*, a_2^*, z; \theta)|\Omega_1\right] - E\left[r_1(a, a_2^*, z; \theta)|\Omega_1\right] > 0, \text{ for } a \neq a_1^* \\ E\left[r_2(a_1^*, a_2^*, z; \theta)|\Omega_2\right] - E\left[r_2(a_1^*, a, z; \theta)|\Omega_2\right] > 0, \text{ for } a \neq a_2^* \end{aligned} \tag{3}$$

To operationalize this, consider some instruments $Z_{1i}, \ldots, Z_{Mi} \in \Omega_i$, and transform them such that they are non-negative-valued. Then, the conditional moment inequalities above imply the unconditional inequalities

$$\begin{aligned} E\left[(r_1(a_1^*, a_2^*, z; \theta) - r_1(a, a_2^*, z; \theta)) * Z_m\right] > 0, \text{ for } a \neq a_1^*,\ m = 1, \ldots, M \\ E\left[(r_2(a_1^*, a_2^*, z; \theta) - r_2(a_1^*, a, z; \theta)) * Z_m\right] > 0, \text{ for } a \neq a_2^*,\ m = 1, \ldots, M. \end{aligned} \tag{4}$$

Accordingly, these unconditional moments can be estimated by sample averages.

Example: Ho, Ho, and Mortimer (2012)

[2] We can relax this to allow for $\pi_i(a_1, a_2, z) = r_i(a_1, a_2, z; \theta) + v_{i,a_1,a_2}$, where $v_{\ldots}$ are errors which have mean zero conditional on Ω_i.

2 Inference procedures with moment inequalities/incomplete models

In cases when a model is not sufficient to point-identify a parameter θ, goal of estimation is to recover the "identified set": the set of θ's (call this Θ_0) which satisfy population analogs of moment inequalities $Eg(x,\theta) \geq 0$:

$$\Theta_0 = \{\theta : \; Eg(x,\theta) \geq 0\}.$$

With small samples, we will never know Θ_0 exactly.

2.1 Identified parameter vs. identified set

The existing literature stresses two approaches for inference in partially identified models: deriving confidence sets with either cover the (i) identified set or (ii) the elements in the identified set with some prescribed probability. More formally, a given confidence set $\hat{\Theta}_n$ satisfies either of the asymptotic conditions

$$\text{liminf}_{n\to\infty}\text{inf}_{\theta\in\Theta_0} P(\theta \in \hat{\Theta}_n) = 1-\alpha; \qquad \text{or}$$

$$\text{liminf}_{n\to\infty} P(\Theta_0 \subset \hat{\Theta}_n) = 1-\alpha;$$

where $1-\alpha$ denotes a prescribed coverage probability.

This distinction was emphasized by Imbens and Manski (2004). Generally, CS for "identified parameter" will be smaller than CS for "identified set". Intuition: consider identified interval $\theta \in [a,b]$ with estimators $\hat{a}_n$ and $\hat{b}_n$. For asymptotic normal estimates, we can form the symmetric two-sided $1-\alpha$ confidence interval as $[\underline{a}_n, \bar{a}_n] = \hat{a}_n \pm z_{\alpha/2}\sigma/\sqrt{n}$, where $z_{1-\alpha/2}$ denotes $(1-\alpha/2)$-th quantile of $N(0,1)$; analogously for $[\underline{b}_n, \bar{b}_n]$.

Consider the intuitive confidence region $\mathcal{C}_n \equiv [\underline{a}_n, \bar{b}_n]$. This should cover identified interval $[a,b]$ with asymptotic probability $(1-\alpha)$, for the usual reasons.

But consider coverage probability of $\mathcal{C}_n$ for any point $\theta \in [a,b]$. As $n \to \infty$, any $\theta \in (a,b)$ will lie in $\mathcal{C}$ with probability 1. For $\theta = a$: note that by construction, $[\underline{a}_n, \bar{a}_n]$ covers a with asymptotic probability $(1-\alpha)$, so that $\mathcal{C}_n \supset [\underline{a}_n, \bar{a}_n]$ covers a with probability $\geq (1-\alpha)$ asymptotically. Similarly with $\theta = b$. Indeed, for this case, Imbens and Manski show that the "doubly one-sided interval" $[\hat{a}_n - z_{1-\alpha}\sigma/\sqrt{n}, \hat{b}_n + z_{1-\alpha}\sigma/\sqrt{n}]$ covers each $\theta \in [a,b]$ with asymptotic probability no smaller than $(1-\alpha)$.

2.2 Confidence sets which cover "identified parameters"

Typical approach to confidence set construction: "invert" test of the point hypotheses $H_0 : \theta \in \Theta_0 \Leftrightarrow Eg(x, \theta) \geq 0$ vs. $H_1 : \theta \notin \Theta_0$. Different confidence sets arise from using different test statistics (Wald-type statistics, empirical likelihood statistics, etc.) For a given test statistic, let the critical value $c_{1-\alpha}(\theta)$ denote the (asymptotic) $1 - \alpha$ quantile, under the null. Then we form our confidence set:

$$\hat{\Theta}_n = \{\theta : \ T_n(\theta) \leq c_{1-\alpha}(\theta)\}$$

where $T_n(\theta)$ denotes the sample test statistic, evaluated at the parameter θ.

Asymptotic distributions of these test statistics are typically derived and known for the case of moment *equalities*: $Eg(x, \theta) = 0$. When we have moment inequalities, then these test statistics take non-standard forms, usually without closed forms. (See Wolak (1989) for early work on deriving these non-standard distributions.) Hence, the critical values for these tests must typically be simulated via resampling procedures (such as bootstrap). Computationally demanding. But for special class of models (which includes entry models and other discrete game models typically estimated in empirical IO), there can be simplifications.

Here we consider the simple two-step framework of Shi and Shum (2012). The model considered consists of two stages. In the first stage, a parameter $\beta \in \mathcal{B} \subset R^{d_\beta}$ is point identified and has a consistent and asymptotically normal (CAN) estimator $\hat{\beta}_n$. In the second stage, the model relates the true value β_0 of β to a structural parameter θ (with true value θ_0), through some inequality/equality restrictions:

$$\begin{aligned} g^e(\theta_0, \beta_0) &= 0 \\ g^{ie}(\theta_0) &\geq 0, \end{aligned} \tag{5}$$

where $g^{ie} : \Theta \to R^{d_1}$ defines the the inequality restrictions, $g^e : \Theta \times \mathcal{A} \to R^{d_2}$ defines the equality restrictions and $\theta \in \Theta \subset R^{d_\theta}$. The parameter θ is potentially partially-identified. The identified set of θ is

$$\Theta_0 = \{\theta \in \Theta : g^e(\theta, \beta_0) = 0 \text{ and } g^{ie}(\theta) \geq 0\}. \tag{6}$$

Note that the only source of sampling error in this framework derives from the estimation of $\hat{\beta}$; also, the inequality constraints do not contain β, and hence are *completely deterministic*.[3]

[3] This is not particularly restrictive because one can always convert an inequality constraint into an equality constraint by introducing a slackness parameter, say γ, and adding an inequality constraint: $\gamma \geq 0$. See entry example below.

(Example: Entry Game) This follows Andrews, Berry, and Jia (2004) and Ciliberto and Tamer (2009). Player j, $j = 1, 2$ enters the market if the profit of entering exceeds 0: $y_j = \{\pi_j \geq 0\}$. The profit $\pi_j = a_j + \delta_j y_{-j} + \varepsilon_j$, where a_j is the expected monopoly profit, δ_j is the competition effect which is assumed to be negative and $(\varepsilon_1, \varepsilon_2)$ follows a distribution known up to a parameter σ: $F(\cdot, \cdot; \sigma)$. Then the model predicts the probabilities of $(0, 0)$ and $(1, 1)$: $g_{00}(a, \delta, \sigma)$ and $g_{11}(a, \delta, \sigma)$ and the upper bounds for the probabilities of $(0, 1)$ and $(1, 0)$: $g_{01}(a, \delta, \sigma)$ and $g_{10}(a, \delta, \sigma)$, where $a = (a_1, a_2)'$ and $\delta = (\delta_1, \delta_2)'$. The outcome probabilities $p_{00}, p_{11}, p_{01}, p_{10}$ are the first stage point identified parameters. In the second stage, the structural parameters (a, δ, σ) are identified by the equalities/inequalities:

$$\begin{aligned} g_{00}(a, \delta, \sigma) - p_{00} &= 0 \\ g_{11}(a, \delta, \sigma) - p_{11} &= 0 \\ g_{01}(a, \delta, \sigma) - p_{01} &\geq 0 \\ g_{10}(a, \delta, \sigma) - p_{10} &\geq 0. \end{aligned} \tag{7}$$

The equalities/inequalities in (7) do not fall immediately into our general framework because the inequalities involve the first-stage parameters. However, we can introduce a nuisance second stage parameter γ, add the restriction $\gamma = p_{01}$ and rewrite the inequalities to only involve $(a, \delta, \sigma, \gamma)$. Specifically, let $\beta = (p_{00}, p_{11}, p_{01}, p_{10})$, $\theta = (a, \delta, \sigma, \gamma)$ for a nuisance parameter $\gamma \in [0, 1]$,

$$\begin{aligned} g^e(\theta, \beta) &= \begin{pmatrix} g_{00}(a, \delta, \sigma) - p_{00} \\ g_{11}(a, \delta, \sigma) - p_{11} \\ \gamma - p_{01} \end{pmatrix}, \text{ and} \\ g^{ie}(\theta) &= \begin{pmatrix} g_{01}(a, \delta, \sigma) - \gamma \\ g_{10}(a, \delta, \sigma) - (1 - g_{00}(a, \delta, \sigma) - g_{11}(a, \delta, \sigma) - \gamma) \end{pmatrix}. \end{aligned} \tag{8}$$

Then the entry game model is written in the form of (5).[4] ■

(Example: Discrete mixture model) Consider a structural model with discrete unobserved heterogeneity, where a (discrete) outcome variable y is drawn according to a known parametric mixture distribution $f(y|\theta, \eta)$ characterized by structural parameters σ and mixing parameter η. Assuming that y takes K distinct values, and η takes M distinct values,

[4] Allowing covariates is easy. We can simply estimate $p_{00}(x) \equiv \Pr(0, 0|x), ..., p_{10}(x) \equiv \Pr(1, 0|x)$ in the first stage either fully nonparametrically, or use some flexible parametric form. Then in the second stage, use $(p_{00}(x), p_{11}(x), p_{01}(x), p_{10}(x))$ in place of $(p_{00}, p_{11}, p_{01}, p_{10})$. The estimated $a(x)$ and $\delta(x)$ will be the monopoly profit and the competition effects conditional on x. Generalizing the example to a game with more than 2 players can be done following Ciliberto and Tamer (2009).

the model is given by the equality constraints

$$P(y=k)=\sum_{m=1}^{M} f(k|\sigma,\eta=m)p_m, \text{ for } k=1,\dots,K; \quad \sum_{m=1}^{M} p_m=1.$$

In this example, the observed probabilities $P(y=k), k=1,\dots,K$ are our β, and $(\sigma,\vec{p}_\eta)$ is our θ where $\vec{p}_\eta=(p_1,...,p_M)'$. Examples of such models are the entry game with multiple equilibria in Bajari, Hahn, Hong, and Ridder (2011) and the structural nonlinear panel data models in Bonhomme (forthcoming).■

Define a criterion function

$$Q(\theta,\beta;W)=g^e(\theta,\beta)'Wg^e(\theta,\beta), \tag{9}$$

where W is a positive definite matrix. Then it clear that

$$\Theta_0=\arg\min_{\theta\in\Theta} Q(\theta,\beta_0;W) \ s.t. \ g^{ie}(\theta)\geq 0. \tag{10}$$

To define the confidence set, we choose a specific weighting matrix $\hat{W}$:

$$\hat{W}^*(\theta)=\left[G(\theta,\hat{\beta})\hat{V}_\beta G(\theta,\hat{\beta})'\right]^{-1}, \tag{11}$$

where $G(\theta,\beta)=\partial g^e(\theta,\beta)/\partial\beta'$, $\hat{V}_\beta$ is a consistent estimator of the asymptotic variance of $\tau_n(\hat{\beta}-\beta)$, where τ_n is a normalizing sequence, e.g. $\tau_n=\sqrt{n}$. Define the confidence set to be

$$CS_n=\{\theta: g^{ie}(\theta)\geq 0, \tau_n^2 Q(\theta,\hat{\beta};\hat{W}^*(\theta))\leq\chi^2_{d_2}(1-\alpha)\}, \tag{12}$$

where $\chi^2_{d_2}(1-\alpha)$ is the $1-\alpha$ quantile of the chi-squared distribution with d_2 degrees of freedom and $1-\alpha\in(0,1)$ is the confidence level.

Theorem 1 *Suppose that $\tau_n(\hat{\beta}-\beta)\to_d Z_\beta\sim N(0,V_\beta)$, $g^e(\theta,\beta)$ is continuously differentiable in β, $G(\theta,\beta)$ is continuous in θ and β, $G(\theta,\beta_0)V_\beta G(\theta,\beta_0)'$ is invertible for all $\theta\in\Theta$ and $\hat{V}_\beta\to_p V_\beta$. Also suppose that $\Theta\times\mathcal{B}$ is compact and g^e and g^{ie} are continuous in Θ. Then*

(a) $\liminf_{n\to\infty}\inf_{\theta\in\Theta_0}\Pr(\theta\in CS_n)=\limsup_{n\to\infty}\sup_{\theta\in\Theta_0}\Pr(\theta\in CS_n)=1-\alpha$*;*

(b) in addition, the following condition (***) holds

$$G(\theta_1,\beta_0)=G(\theta_2,\beta_0) \quad \textit{for all } \theta_1,\theta_2\in\Theta_0 \tag{***}$$

then $\lim_{n\to\infty}\Pr(\Theta_0\subseteq CS_n)=1-\alpha$.

The additional assumption (***) for part (b) is immediately satisfied if θ and β are additively separable in g^e, as they are in all the previous examples. Additive separability is likely to hold in models in which the equality restrictions take the form of "matching" empirical frequencies to outcome probabilities predicted by the model, which is a common feature of all the examples above.

[Example: Iaryczower, Shi, and Shum (2012)]

2.3 Confidence sets which cover the identified set

We discuss here the estimation procedure of Romano and Shaikh (2010). Let P denote true (but unknown) data-generating process of data, and let Θ denote the parameter space. Define identified set as:

$$\Theta_0(P) \equiv \text{argmin}_{\theta\in\Theta} Q(\theta, P) \Leftrightarrow \{\theta \in \Theta : \; Q(\theta, P) = 0\}.$$

Let $Q_n(\theta)$ denote the sample objective function, and a_n a rate of convergence such that as $n \to \infty$, $a_n \to \infty$ and $a_n Q_n(\theta) \stackrel{d}{\to} \mathcal{L}(\theta)$, a nondegenerate limiting distribution.

For the case of inequality constrained moment conditions $E_P g_m(Y,\theta) \leq 0$, for m indexing the moment conditions, one possibility is to define a least-squares objective function $Q(\theta, P) = \sum_m [E_p g_m(Y,\theta)]_+^2$ with small-sample analog $Q_n(\theta) = \sum_m [\frac{1}{T} g_m(Y_t,\theta)]_+^2$.

The notation $[y]_+$ is shorthand for $y \cdot \mathbf{1}(y > 0)$. For the specific case of the first inequality in the conditions above, $g(Y_t, \theta) = Y_{t1} - [1 - \Phi(-\beta' s)][\Phi(\delta - \beta' s)]$.

Goal of estimation: recover confidence set C_n s.t.

$$\text{liminf}_{n\to\infty} P\{C_n \supseteq \Theta_0(P)\} = 1 - \alpha \tag{13}$$

for some level α.

Romano-Shaikh show:

- This estimation problem is equivalent to a testing problem, where the goal is to test the family (continuum) of hypotheses

  $$H_\theta : \; \theta \in \Theta_0(P), \; \forall\, \theta \in \Theta$$

  subject to a restriction on the "family-wise error rate" FWER such that

  $$\text{limsup}_{n\to\infty} FWER_{P,n} = \alpha \tag{14}$$

where

$$FWER_{P,n} = P\left\{\text{with } n \text{ obs., reject at least 1 ``true'' null hypothesis } H_\theta \text{ s.t. } Q(\theta, P) = 0\right\}.$$

FWER is a generalization of type I error (when you have uni-dimensional hypothesis test).

This does not define the test statistic, just a characteristic that the test should satisfy.

- Note that

$$\begin{aligned} 1 - FWER_{P,n} &= P\left\{\text{with } n \text{ obs., reject no null hypothesis } H_\theta \text{ s.t. } Q(\theta, P) = 0\right\} \\ &= P\left\{\Theta_0(P) \subseteq C_n\right\} \end{aligned} \tag{15}$$

Hence the FWER restriction (14) is equivalent to

$$\begin{aligned} &\text{limsup}_{n\to\infty} 1 - P\left\{\Theta_0(P) \subseteq C_n\right\} = \alpha \\ \Leftrightarrow &\text{liminf}_{n\to\infty} P\left\{\Theta_0(P) \subseteq C_n\right\} = 1 - \alpha \end{aligned} \tag{16}$$

which is the required criterion (13).

- The following "stepdown" algorithm yields, at the end, a set C_n satisfying Eq. (13).

1. Start with $S^1 = \Theta$ (the entire parameter space).
2. Evaluate this test statistic:

$$\tau_n(S^1, \theta) \equiv \max_{\theta \in S^1} a_n Q_n(\theta).$$

3. Compare test statistic $\tau_n(S^1, \theta)$ to critical value $c_n(S^1, 1-\alpha)$, where critical value is obtained by subsampling:
   - Consider subsamples of size b_n from original dataset. There are $N_n \equiv \binom{n}{b_n}$ subsampled datasets from the original dataset, indexed by $i = 1, \ldots, N_n$. (As $n \to \infty$, $\frac{b_n}{n} \to 0$.)
   - For each subsampled dataset i: compute the subsampled test statistic

$$\kappa_{i,n}(S^1, \theta) \equiv \max_{\theta \in S^1} a_b Q_b(\theta).$$

   - Set $c_n(S^1, 1-\alpha)$ to $(1-\alpha)$-th quantile amongst $\left\{\kappa_{1,n}(S^1, \theta), \ldots, \kappa_{N_n,n}(S^1, \theta)\right\}$.
4. If $\tau_n(S^1, \theta) \leq c_n(S^1, 1-\alpha)$, then stop, and set $C_n = S^1$,
5. If $\tau_n(S^1, \theta) > c_n(S^1, 1-\alpha)$, then set

$$S^2 = \left\{\theta \in S^1 : a_n Q_n(\theta) \leq c_n(S^1, 1-\alpha)\right\}.$$

   Repeat from Step 2, using S^2 in place of S^1.

3 Random set approach

- Beresteanu and Molinari (2008), Beresteanu, Molchanov, and Molinari (2011)
- Probability space $(\Omega, \mathcal{A}, \mu)$
- Set-valued random variable ("random set" for short) $F: \Omega \mapsto K(\mathbb{R}^d)$, where $K(\mathbb{R}^d)$ denotes the set of nonempty closed subsets in $\mathbb{R}^d$.
- Inner product $< \cdot, \cdot >$ in $\mathbb{R}^d$ is vector dot product; Euclidean distance between $\vec{a}, \vec{b} \in \mathbb{R}^d$ is $||\vec{a} - \vec{b}|| = \sqrt{< \vec{a} - \vec{b}, \vec{a} - \vec{b} >}$.
- $K(\mathbb{R}^d)$ is metric space with Hausdorff distance:

$$H(A, B) = \max(d_H(A, B), d_H(B, A)), \quad d_H(A, B) = \sup_{\vec{a} \in A} \inf_{\vec{b} \in B} ||\vec{a} - \vec{b}||.$$

- Selection of random set F is random vector $\vec{f} : \Omega \mapsto \mathbb{R}^d$ such that $\vec{f}(\omega) \in F(\omega)$. $S(F)$ is set of selections of F.
- Aumann expectation: $\mathbb{E}(F) = \left\{ \int_\Omega \vec{f} d\mu, \ \vec{f} \in S(F) \right\}$.

  Note: for given $\vec{f}$, $\int_\Omega \vec{f} d\mu$ is d-dim. vector with i-th element $\int_\Omega f_i(\omega)\mu(d\omega)$.
- Support function for a set $R \in K(\mathbb{R}^d)$ is $s(\vec{p}, R) = \sup_{\vec{r} \in R} < \vec{p}, \vec{r} >$, for all $\vec{p} \in \mathbb{R}^d$. Support function for a set R is the *Fenchel/Legendre transform* of the indicator function for that set, defined as $\delta_R(\vec{p}) = 0 \cdot \mathbf{1}(\vec{p} \in R) + (+\infty) \cdot \mathbf{1}(\vec{p} \notin R)$.[5] In this sense, the support function is an equivalent representation of the set R.
- Minkowski summation: $A \oplus B = \left\{ \vec{a} + \vec{b} : \ \vec{a} \in A, \vec{b} \in B \right\}$.
- Random set limit theorems: for $\{F, F_i, i = 1, 2, \ldots\}$ being i.i.d. random sets, we have
  - Law of large numbers:

$$H\left(\frac{1}{n} \oplus_{i=1}^n F_i, \mathbb{E}(F)\right) \stackrel{as}{\to} 0.$$

  - Central limit theorem:

$$\sqrt{n} H\left(\frac{1}{n} \oplus_{i=1}^n F_i, \mathbb{E}(F)\right) \stackrel{d}{\to} ||z||_{\mathbb{C}(\mathbb{S}^{d-1})}$$

[5] By definition of the Fenchel transform (cf. Borwein and Lewis (2006, pg. 55)), the Fenchel conjugate of $\delta_R(\vec{p})$ is the function $\delta_R^*(\vec{p}) \equiv \sup_{\vec{r} \in R} \{< \vec{p}, \vec{r} > - \delta_R(r)\} = \sup_{\vec{r} \in R} \{< \vec{p}, \vec{r} >\} = s(\vec{p}, R)$.

where $||z||_{\mathbb{C}(\mathbb{S}^{d-1})} = \sup_{\vec{p}\in\mathbb{S}^{d-1}}|z(\vec{p})|$ and $z : \mathbb{S}^{d-1} \mapsto \mathbb{R}$ is a continuous scalar-valued Gaussian random function: $z(\vec{p}) \sim N(0, E(s(\vec{p}, F)^2) - [Es(\vec{p}, F)]^2)$ and $Cov\,(z(\vec{p}), z(\vec{q})) = E[s(\vec{p}, F)s(\vec{q}, F)] - Es(\vec{p}, F) \cdot Es(\vec{q}, F)$.

- Hypothesis testing:

$$H_0 : \ \mathbb{E}(F) = \Psi_0 \quad vs. \quad H_0 : \ \mathbb{E}(F) \neq \Psi_0$$

Define the critical value c_α from $P\left\{||z||_{\mathbb{C}(\mathbb{S}^{d-1})} > c_\alpha\right\}$, where α is the size. This can be simulated.

Then an α-size test is: reject H_0 if

$$\sqrt{n}H\left(\frac{1}{n}\oplus_{i=1}^{n} coF_i, \Psi_0\right) > c_\alpha.$$

- Confidence set can be obtained by inverting this test:

$$CC_{n,1-\alpha} = \left\{\tilde{\Psi} \in K(\mathbb{R}^d) : \ \sqrt{n}H\left(\frac{1}{n}\oplus_{i=1}^{n} coF_i, \tilde{\Psi}\right) \leq c_\alpha\right\}.$$

This is a collection of sets. Let $\mathcal{U}_n \equiv \cup\left\{\tilde{\Psi} \in CC_{n,1-\alpha}\right\}$. Then theorem 2.4 in the paper shows:

$$\mathcal{U}_n = \left(\frac{1}{n}\oplus_{i=1}^{n} coF_i\right) \oplus B_{c_\alpha}, \quad B_{c_\alpha} = \left\{b \in \mathbb{R}^d : \ ||b|| \leq c_\alpha/\sqrt{n}\right\}.$$

3.1 Application: sharp identified region for games with multiple equilibria

- Outcome space $\mathcal{Y} \subset \mathbb{R}^k$; randomness indexes by $\omega \in \Omega$ with probability P.
- Random set $Q(\omega; \theta) \subset \mathcal{Y}$, for all $\omega \in \Omega$: for each ω, the realization of the random set $Q(\omega; \theta)$ is the set (possibly singleton) of outcomes which could occur in an equilibrium of the game, given (ω, θ).
- Aumann expectation $\mathbb{E}Q(\omega; \theta) \subset \mathcal{Y}$: set of expected outcomes consistent with some equlibrium selection rule.
- Average outcomes observed in data: $E(y)$, for $y \in \mathcal{Y}$.

- Sharp identified region:

$$\Theta_0 = \{\theta : E(y) \in \mathbb{E}Q(\omega;\theta)\}$$

which is typically a convex set. Using support function $s(u, \mathbb{E}Q(\omega;\theta))$, we can equivalently denote this by (this is the *separating hyperlane theorem*!)

$$\begin{aligned}\Theta_0 &= \{\theta : E(y) \in \mathbb{E}Q(\omega;\theta)\} \\ &= \left\{\theta : u'E(y) \leq s(u, \mathbb{E}Q(\omega;\theta)),\ \forall u \in \mathbb{R}^k, ||u|| = 1\right\} \\ &= \left\{\theta : u'E(y) \leq Es(u, Q(\omega;\theta)),\ \forall u \in \mathbb{R}^k, ||u|| = 1\right\} \\ &= \left\{\theta : \max_{u \in \mathbb{R}^k, ||u||=1} \left[u'E(y) - Es(u, Q(\omega;\theta))\right] \leq 0\right\}\end{aligned}$$

which, for each θ, is a k-dimensional optimization program. The final equality in the above display uses the random set result that

$$s(u, \mathbb{E}Q(\omega;\theta)) = Es(u, Q(\omega;\theta)).$$

3.2 Application: interval-censored regression

- Consider regression model: $y = \theta_1 + \theta_2 x + u$, with exogenous u.
- Point-identified case: moment based estimation

$$E(y) = \theta_1 + \theta_2 E(x); \quad E(xy) = \theta_1 E(x) + \theta_2 E(x^2)$$

$$\Rightarrow \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & E(x) \\ E(x) & E(x^2) \end{pmatrix}^{-1}}_{\equiv \Sigma} \begin{pmatrix} E(y) \\ E(xy) \end{pmatrix}$$

- Interval censoring model: random vector (y, y_L, y_U, x) with $P(y_L \leq y \leq y_U) = 1$.
- Only observe (y_{iL}, y_{iU}, x_i), $i = 1, \ldots, n$.
- Define the i.i.d. random sets

$$G(\omega) = \left\{ \begin{bmatrix} y(\omega) \\ x(\omega) \cdot y(\omega) \end{bmatrix} : \ y(\omega) \in [y_L(\omega), y_U(\omega)] \right\}. \quad \text{and}$$

$$G_i = \left\{ \begin{bmatrix} y \\ x_i \cdot y \end{bmatrix} : \ y \in [y_{iL}, y_{iU}] \right\} \quad i = 1, \ldots, n.$$

- Define identified set

$$\begin{aligned}\Theta &= \Sigma^{-1}\mathbb{E}(G)\\ &= \left\{ \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \Sigma^{-1} \begin{pmatrix} E(y) \\ E(xy) \end{pmatrix}, \begin{pmatrix} y \\ xy \end{pmatrix} \in S(G) \right\}.\end{aligned}$$

- Sample analog estimator:

$$\hat{\Theta}_n = \hat{\Sigma}^{-1} \cdot \left(\frac{1}{n} \oplus_{i=1}^{n} G_i \right).$$

Consistency straightforward. Confidence set more involved.

References

Andrews, D., S. Berry, and P. Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," mimeo., Yale University.

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

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

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

Bonhomme, S. (forthcoming): "Functional Differencing," *Econometrica.*

Borwein, J., and A. Lewis (2006): *Convex Analysis and Nonlinear Optimization.* Springer Verlag.

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

Galichon, A., and M. Henry (forthcoming): "Inference in models with multiple equilibria," *Review of Economic Studies.*

Ho, K., J. Ho, and J. Mortimer (2012): "The Use of Full-Line Forcing Contracts in the Video Rental Industry," *American Economic Review*.

Iaryczower, M., X. Shi, and M. Shum (2012): "Can Deliberation Trump Conflict? Partial Identification of a Deliberative Voting Model," unpublished manuscript, Princeton University.

Imbens, G., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2007): "Moment Inequalities and their Application," Manuscript, Harvard University.

Romano, J., and A. Shaikh (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–211.

Shi, X., and M. Shum (2012): "Simple Two-stage Inference for a Class of Partially Identified Models," mimeo., UW-Madison.

Tamer, E. (2003): "Incomplete simultaneous discrete response model with multiple equilibria," *Review of Economic Studies*, 70, 147–165.

Wolak, F. (1989): "Local and Global Testing of Linear and Nonlinear Inequality Constraints in Nonlinear Econometric Models," *Econometric Theory*, 1, 1–35.