ECON 481-3
LECTURE 14: INFERENCE IN MOMENT INEQUALITY MODELS

Ivan A. Canay
Northwestern University

1. Canay, I.A. and A.M. Shaikh (2017): "Practical and Theoretical Advances for Inference in Partially Identified Models", In B. Honore, A. Pakes, M. Piazzesi, & L. Samuelson (Eds.), Advances in Economics and Econometrics: Volumen 2: Eleventh World Congress.
2. Ho, K. and A. M. Rosen (2017): "Partial Identification in Applied Research: Benefits and Challenges", In B. Honore, A. Pakes, M. Piazzesi, & L. Samuelson (Eds.), Advances in Economics and Econometrics: Volumen 2: Eleventh World Congress.

Outline of Lecture

- Examples leading to moment inequalities
  - – Entry Games
  - – Revealed Preferences in Discrete Choice
  - – Missing data
- Confidence regions for partially identified models
  - – Importance of uniform asymptotic validity
- Moment inequalities: five distinct approaches
  1. Least Favorable Test
  2. subsampling
  3. Moment Selection
  4. Refined Moment Selection
  5. Two-step methods
- Subvector inference for moment inequalities (Skip today)
- Extensions

Notation

Some basic notation:

$\hat{P}_n =$ empirical distribution of $W_i, i = 1, \ldots, n$.

$\mu(\theta, P) = E_P[m(W_i, \theta)]$.

$\bar{m}_n(\theta) =$ sample mean of $m(W_i, \theta)$.

$\hat{\Omega}_n(\theta) =$ sample correlation of $m(W_i, \theta)$.

$\sigma_j^2(\theta, P) = \text{Var}_P[m_j(W_i, \theta)]$.

$\hat{\sigma}_{n,j}^2(\theta) =$ sample variance of $m_j(W_i, \theta)$.

$\hat{D}_n(\theta) = \text{diag}(\hat{\sigma}_{n,1}(\theta), \ldots, \hat{\sigma}_{n,k}(\theta))$.

Recap: The auxiliary distribution

Test Statistic

$$T_n(\theta) = T\left(\hat{D}_n^{-1}(\theta)\sqrt{n}\bar{m}_n(\theta), \hat{\Omega}_n(\theta)\right) .$$

Auxiliary distribution

$$J_n(x, s(\theta), \theta, P) = P\left\{T\left(\hat{D}_n^{-1}(\theta)Z_n(\theta) + \hat{D}_n^{-1}(\theta)s(\theta), \hat{\Omega}_n(\theta)\right) \leqslant x\right\} ,$$

Consider the following derivation:

Recap: Least Favorable and SS

$$P\{T_n(\theta) \leqslant x\} = J_n(x, \sqrt{n}\mu(\theta, P), \theta, P)$$

is **hard** to estimate due to presence of $\sqrt{n}\mu(\theta, P)$, where

$$J_n(x, s(\theta), \theta, P) = P\left\{T\left(\hat{D}_n^{-1}(\theta)Z_n(\theta) + \hat{D}_n^{-1}(\theta)s(\theta), \hat{\Omega}_n(\theta)\right) \leqslant x\right\} ,$$

and

$$Z_n(\theta) = \sqrt{n}(\bar{m}_n(\theta) - \mu(\theta, P)) .$$

However, $J_n(x, s(\theta), \theta, P)$ is easy to estimate for a given function $s(\theta)$.

- Least Favorable: $\sqrt{n}\mu(\theta, P) \leqslant 0$ for any $P \in \mathbf{P}$ and $\theta \in \Theta_0(P)$

$$\implies J_n^{-1}(1-\alpha, \sqrt{n}\mu(\theta, P), \theta, P) \leqslant J_n^{-1}(1-\alpha, 0, \theta, P) .$$

- Subsampling: implicitly uses $\sqrt{n}\mu(\theta, P) \leqslant \sqrt{b}\mu(\theta, P)$ as upper bound.

$$\implies J_n^{-1}(1-\alpha, \sqrt{n}\mu(\theta, P), \theta, P) \leqslant J_n^{-1}(1-\alpha, \sqrt{b}\mu(\theta, P), \theta, P) .$$

Generalized Moment Selection

- **Main Idea**: Perhaps possible to estimate $\sqrt{n}\mu(\theta, P)$ "well enough"?

- **Selection fn:** consider, e.g., $\hat{s}_n^{\text{gms}}(\theta) = (\hat{s}_{n,1}^{\text{gms}}(\theta), \ldots, \hat{s}_{n,k}^{\text{gms}}(\theta))'$ with

$$\hat{s}_{n,j}^{\text{gms}}(\theta) = \begin{cases} 0 & \text{if } \frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} > -\kappa_n \\ -\infty & \text{otherwise} \end{cases},$$

  where $0 < \kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$.

- Choosing

$$\hat{c}_n(1-\alpha, \theta) = \text{ estimate of } J_n^{-1}\left(1-\alpha, \hat{s}_n^{\text{gms}}(\theta), \theta, P\right)$$

  leads to valid tests.

See Andrews & Soares (2010). Related results in Bugni (2010) and Canay (2010).

Generalized Moment Selection (cont.)

GMS

The GMS test takes the form

$$\phi_n^{\text{gms}}(\theta) \equiv I\left\{T_n(\theta) > \widehat{J}_n^{-1}(1-\alpha, \hat{s}_n^{\text{gms}}(\theta), \theta)\right\}.$$

- **Why does it work?** note that

$$\frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} = \frac{\sqrt{n}(\bar{m}_{n,j}(\theta) - \mu_j(\theta, P))}{\hat{\sigma}_{n,j}(\theta)} + \frac{\sqrt{n}\mu_j(\theta, P)}{\hat{\sigma}_{n,j}(\theta)}.$$

  - First term: $O_P(1)$ for θ and P s.t. $\mu_j(\theta, P) \leqslant 0$.
  - Second term: either is zero or diverges in probability to $-\infty$ depending, respectively, on whether $\mu_j(\theta, P) = 0$ or $\mu_j(\theta, P) < 0$.

Generalized Moment Selection (cont.)

$$\frac{1}{\kappa_n}\frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} = \frac{1}{\kappa_n}\frac{\sqrt{n}(\bar{m}_{n,j}(\theta) - \mu_j(\theta, P))}{\hat{\sigma}_{n,j}(\theta)} + \frac{1}{\kappa_n}\frac{\sqrt{n}\mu_j(\theta, P)}{\hat{\sigma}_{n,j}(\theta)} .$$

$$\hat{s}^{\text{gms}}_{n,j}(\theta) = \begin{cases} 0 & \text{if } \frac{1}{\kappa_n}\frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} > -1 \\ -\infty & \text{otherwise} \end{cases} ,$$

- It follows that for any sequence $P_n \in \mathbf{P}$ and $\theta_n \in \Theta_0(P_n)$

$$\hat{s}^{\text{gms}}_{n,j}(\theta_n) = \begin{cases} 0 & \text{if } \sqrt{n}\mu_j(\theta_n, P_n) \to c \leqslant 0 \\ -\infty & \text{if } \sqrt{n}\mu_j(\theta_n, P_n) \to -\infty \end{cases} \quad \text{w.p.a.1} .$$

  In this sense, $\hat{s}^{\text{gms}}_n(\theta)$ provides an asymptotic upper bound on $\sqrt{n}\mu(\theta, P)$.

- **Alternative Interpretation**: $\hat{s}^{\text{gms}}_{n,j}(\theta)$ "selects" whether $\mu_j(\theta, P) = 0$ or $\mu_j(\theta, P) < 0$.

Generalized Moment Selection (cont.)

- **Remark**: as SS, it incorporates information about $\sqrt{n}\mu(\theta, P)$ ...

  ... and, for typical κ_n and b, **more powerful** than subsampling.

- Main drawback is **choice of κ_n**:

  – In finite-samples, smaller choice always more powerful.

  – First- and higher-order properties do not depend on κ_n.

  See Bugni (2014).

  – Precludes data-dependent rules for choosing κ_n.

- **Power**: tests use the same $T_n(\theta)$ so power comparison only entail comparisons of critical values

- **Intuition**: the method that detects non-binding moments more effectively will lead to a test with higher power.

Asymptotic Power - Intuition

$$\hat{s}^{\text{gms}}_{n,j}(\theta) = \begin{cases} 0 & \text{if } \kappa_n^{-1}\sqrt{n}\bar{m}_{n,j}(\theta) > -1 \\ -\infty & \text{otherwise} \end{cases} \quad (\text{no } \hat{\sigma}_{n,j}(\theta))$$

Example

Consider a model with two moments and sequences θ_n and P_n such that $\sigma_j^2(\theta, P) = 1$ and

$$\mu_1(\theta_n, P_n) = \frac{h_1}{\sqrt{n}} < 0 \quad \text{and} \quad \mu_2(\theta_n, P_n) = \frac{\pi_1}{\kappa_n^{-1}\sqrt{n}} \in (-\infty, -1) \ .$$

- GMS: first moment treated as binding,
- GMS: second moment treated as **slack**,

Asymptotic Power - Intuition

Consider the following assumption to study SS:

$$\kappa_n \sqrt{\frac{b}{n}} \to 0\,. \qquad (\clubsuit)$$

Example (cont.)

- SS: first moment treated as binding,
- SS: second moment treated as binding,

($\clubsuit$) holds for typical choices $\kappa_n \approx \log n$ and $b_n \approx n^a$ for $a \in (0, 1)$.

QUESTIONS?

Refined Moment Selection

- First/second-order asymptotic properties of GMS tests **do not depend** on κ_n.
- **Finite samples**: a smaller choice of κ_n translate into better power.
- **Main Idea**: In order to develop data-dependent rules for choosing κ_n, ...

  ... change asymptotic framework so κ_n does not depend on n.
- Consider, e.g., $\hat{s}_n^{\text{rms}}(\theta) = (\hat{s}_{n,1}^{\text{rms}}(\theta), \ldots, \hat{s}_{n,k}^{\text{rms}}(\theta))'$ with

$$\hat{s}_{n,j}^{\text{rms}}(\theta) = \begin{cases} 0 & \text{if } \frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} > -\kappa \\ -\infty & \text{otherwise} \end{cases} .$$

- **Note**: $\hat{s}_n^{\text{rms}}(\theta)$ no longer an asymptotic upper bound on $\sqrt{n}\mu(\theta, P)$, so ...

  ... critical value replacing $\hat{s}_n^{\text{gms}}(\theta)$ with $\hat{s}_n^{\text{rms}}(\theta)$ is **too small**.
- For an appropriate size-correction factor $\hat{\eta}_n(\theta) > 0$, choosing

$$\hat{c}_n(1-\alpha, \theta) = \text{ estimate of } J_n^{-1}(1-\alpha, \hat{s}_n^{\text{rms}}(\theta), \theta, P) + \hat{\eta}_n(\theta)$$

  leads to valid tests (whose first-order properties depend on κ.)

Refined Moment Selection

Refined moment selection tests are tests of the form

$$\phi_n^{\text{rms}}(\theta) \equiv I\{T_n(\theta) > \widehat{J}_n^{-1}(1-\alpha, \hat{s}_n^{\text{rms}}(\theta), \theta) + \hat{\eta}_n(\theta)\},$$

where $\hat{\eta}_n(\theta)$ is a size-correction factor.

- In order to determine an appropriate size-correction factor, consider the test

$$\tilde{\phi}_n^{\text{rms}}(\theta) \equiv I\{T_n(\theta) > \hat{J}_n^{-1}(1-\alpha, \hat{s}_n^{\text{rms}}(\theta), \theta) + \eta\}$$

  for an arbitrary non-negative constant η.

- Arguing as before, the limiting rejection probability of this test is

$$P\{T(Z+s^*, \Omega^*) > J^{-1}(1-\alpha, s^{\text{rms},*}(Z+s^*), \Omega^*) + \eta\}.$$

  The appropriate size-correction factor is thus

$$\eta^*(\Omega^*, \kappa) \equiv \inf\left\{\eta > 0 : \sup_{s^* \in \mathbf{R}^k : s^* \leqslant 0} P\{T(Z+s^*, \Omega^*) > J^{-1}(1-\alpha, s^{\text{rms},*}(Z+s^*, \kappa), \Omega^*) + \eta\} \leqslant \alpha\right\}.$$

Refined Moment Selection

Remark: Incorporates information about $\sqrt{n}\mu(\theta, P)$ ...

... in asymptotic framework where first-order properties depend on κ.

Main drawback is **computation of $\hat{\eta}_n(\theta)$**:

- Requires approximate maximum rejection probability over k-dimensional space.
- Andrews & Barwick (2012) simplify the problem in **two ways**:
  - replace Ω by the smallest off diagonal element (δ)
  - examine $2^k - 1$ extreme points, i.e., $s^* \in \{-\infty, 0\}^k$
- Provide numerical evidence in favor of this simplification.
- More results in McCloskey (2015). Still, remains computationally infeasible for $k > 10$.

Precludes many applications, e.g.,

- Bajari, Benkard & Levin (2007) ($k \approx 500$ or more!)
- Ciliberto & Tamer (2009) ($k = 2^{m+1}$ where $m =$ # of firms).

Andrews and Barwick: Table

TABLE I

MOMENT SELECTION TUNING PARAMETERS $\kappa(\delta)$ AND SIZE-CORRECTION FACTORS $\eta_1(\delta)$ AND $\eta_2(p)$ FOR $\alpha = .05$[a]

δ	$\kappa(\delta)$	$\eta_1(\delta)$	δ	$\kappa(\delta)$	$\eta_1(\delta)$	δ	$\kappa(\delta)$	$\eta_1(\delta)$
[−1, −.975)	2.9	.025	[−.30, −.25)	2.1	.111	[.45, .50)	0.8	.023
[−.975, −.95)	2.9	.026	[−.25, −.20)	2.1	.082	[.50, .55)	0.6	.033
[−.95, −.90)	2.9	.021	[−.20, −.15)	2.0	.083	[.55, .60)	0.6	.013
[−.90, −.85)	2.8	.027	[−.15, −.10)	2.0	.074	[.60, .65)	0.4	.016
[−.85, −.80)	2.7	.062	[−.10, −.05)	1.9	.082	[.65, .70)	0.4	.000
[−.80, −.75)	2.6	.104	[−.05, .00)	1.8	.075	[.70, .75)	0.2	.003
[−.75, −.70)	2.6	.103	[.00, .05)	1.5	.114	[.75, .80)	0.0	.002
[−.70, −.65)	2.5	.131	[.05, .10)	1.4	.112	[.80, .85)	0.0	.000
[−.65, −.60)	2.5	.122	[.10, .15)	1.4	.083	[.85, .90)	0.0	.000
[−.60, −.55)	2.5	.113	[.15, .20)	1.3	.089	[.90, .95)	0.0	.000
[−.55, −.50)	2.5	.104	[.20, .25)	1.3	.058	[.95, .975)	0.0	.000
[−.50, −.45)	2.4	.124	[.25, .30)	1.2	.055	[.975, .99)	0.0	.000
[−.45, −.40)	2.2	.158	[.30, .35)	1.1	.044	[.99, 1.0]	0.0	.000
[−.40, −.35)	2.2	.133	[.35, .40)	1.0	.040			
[−.35, −.30)	2.1	.138	[.40, .45)	0.8	.051			

p	2	3	4	5	6	7	8	9	10
$\eta_2(p)$	.00	.15	.17	.24	.31	.33	.37	.45	.50

[a]The values in this table are obtained by simulating asymptotic formulae using 40,000 critical-value and 40,000 rejection-probability simulation repetitions; see Section S7.5 of the Supplemental Material for details.

QUESTIONS?

Two Step Methods

Main Idea:

- **First step:** construct a confidence region for $\mu(\theta, P)$ at some **small** significance level β.
- **Second step:** use this set to provide information about which components of $\mu(\theta, P)$ are "negative" when constructing the test.

Step 1

Construct confidence region $M_n(1-\beta, \theta)$ for $\sqrt{n}\mu(\theta, P)$, s.t.

$$\liminf_{n\to\infty} \inf_{P\in\mathbf{P}} \inf_{\theta\in\Theta_0(P)} P\left\{\sqrt{n}\mu(\theta, P) \in M_n(1-\beta, \theta)\right\} \geqslant 1-\beta \ ,$$

where $0 < \beta < \alpha$.
An upper-right rectangular confidence region is computationally attractive, i.e.,

$$M_n(1-\beta, \theta) = \left\{\mu \in \mathbf{R}^k : \mu_j \leqslant \bar{m}_{n,j}(\theta) + \frac{\hat{\sigma}_{n,j}(\theta)\hat{q}_n(1-\beta, \theta)}{\sqrt{n}}\right\} \ ,$$

where $\hat{q}_n(1-\beta, \theta)$ may be easily constructed using, e.g., bootstrap.

STEP 2

- Use $M_n(1-\beta,\theta)$ to restrict possible values for $\sqrt{n}\mu(\theta,P)$.

  Consider "largest" $s \leqslant 0$ with $s \in M_n(1-\beta,\theta)$, i.e.,

  $$\hat{s}_n^{\text{ts}}(\theta) = (\hat{s}_{n,1}^{\text{ts}}(\theta),\ldots,\hat{s}_{n,k}^{\text{ts}}(\theta))'$$

  with

  $$\hat{s}_{n,j}^{\text{ts}}(\theta) = \min\left\{\sqrt{n}\bar{m}_{n,j}(\theta) + \hat{\sigma}_{n,j}(\theta)\hat{q}_n(1-\beta,\theta),0\right\} .$$

- Choosing

  $$\hat{c}_n(1-\alpha,\theta) = \text{ estimate of } J_n^{-1}(1-\alpha+\beta,\hat{s}_n^{\text{ts}}(\theta),\theta,P) ,$$

  leads to valid tests (whose first-order properties depend on β).

- **Closed-form** expression for $\hat{s}_n^{\text{ts}}(\theta)$ a key feature!

Two Step Methods

Two Step Test

Two-step tests are tests of the form

$$\phi_n^{\text{ts}}(\theta) \equiv I\{T_n(\theta) > \hat{J}_n^{-1}(1-\alpha+\beta, \hat{s}_n^{\text{ts}}, \theta)\} .$$

- The asymptotic validity of these tests relies on

$$P\Big\{T_n(\theta) > \hat{J}_n^{-1}(1-\alpha+\beta, \hat{s}_n^{\text{ts}}(\theta), \theta)\Big\} \leqslant P\Big\{T_n(\theta) > \hat{J}_n^{-1}(1-\alpha+\beta, \sqrt{n}\mu(\theta, P), \theta)\Big\} + P\Big\{\sqrt{n}\mu(\theta, P) \notin M_n(\theta, 1-\beta)\Big\} .$$

- It is straightforward to show that

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{\theta\in\Theta_0(P)} P\Big\{T_n(\theta) > \hat{J}_n^{-1}(1-\alpha+\beta, \sqrt{n}\mu(\theta, P), \theta)\Big\} \leqslant \alpha - \beta .$$

- In addition,

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{\theta\in\Theta_0(P)} P\Big\{\sqrt{n}\mu(\theta, P) \notin M_n(\theta, 1-\beta)\Big\} \leqslant \beta .$$

Two Step Methods

- **Note**: Argument hinges on simple Bonferroni-type inequality.
- **Remark**: Also incorporates information about $\sqrt{n}\mu(\theta, P)$ ...

  ... in asymptotic framework where first-order properties depend on β.
- But, importantly:
  - Remains feasible even for large values of k.
  - Despite "crudeness" of inequality, remains competitive in terms of power.
- Many earlier antecedents:
  - In statistics, e.g., Berger & Boos (1994) and Silvapulle (1996).
  - In economics, e.g., Stock & Staiger (1997) and McCloskey (2012).
  - **Computational simplicity** key novelty here.

QUESTIONS?

Outline of Lecture

- Examples leading to moment inequalities
  - Entry Games
  - Revealed Preferences in Discrete Choice
  - Missing data
- Confidence regions for partially identified models
  - Importance of uniform asymptotic validity
- Moment inequalities: five distinct approaches
  1. Least Favorable Test
  2. subsampling
  3. Moment Selection
  4. Refined Moment Selection
  5. Two-step methods
- Subvector inference for moment inequalities (**Skip**)
- Extensions

Subvector Inference

- Despite advances, methods not commonly employed.
- Methods difficult (infeasible?) when $\dim(\theta)$ even moderately large ...

  ... but interest often only in few coord. of θ (or a fcn. of θ)!
- Let $\lambda(\cdot) : \Theta \to \Lambda$ be function of θ of interest.
- Identified set for $\lambda(\theta)$ is
$$\Lambda_0(P) = \lambda(\Theta_0(P)) = \{\lambda(\theta) : \theta \in \Theta_0(P)\} ,$$
  where
$$\Theta_0(P) = \{\theta \in \Theta : E_P[m(W_i, \theta)] \leqslant 0\} .$$
- **Goal**: Conf. reg. for points in id. set that are unif. consistent in level.
- **Remark**: Methods require same assumptions plus possibly others.

Subvectors: main idea

- **How**: Construct tests $\phi_n(\lambda)$ of

$$H_\lambda : \exists\, \theta \in \Theta \text{ with } E_P[m(W_i,\theta)] \leqslant 0 \text{ and } \lambda(\theta) = \lambda$$

that provide unif. asym. control of Type I error, i.e.,

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}} \sup_{\lambda\in\Lambda_0(P)} E_P[\phi_n(\lambda)] \leqslant \alpha\,.$$

- Given such $\phi_n(\lambda)$,

$$C_n = \{\lambda \in \Lambda : \phi_n(\lambda) = 0\}$$

satisfies desired coverage property.

- Below describe three different tests.

Subvectors: projections

- **Main Idea**: Utilize previous tests $\phi_n(\theta)$:

$$\phi_n^{\text{proj}}(\lambda) = \inf_{\theta \in \Theta_\lambda} \phi_n(\theta) ,$$

where

$$\Theta_\lambda = \{\theta \in \Theta : \lambda(\theta) = \lambda\} .$$

- Properties of $\phi_n(\theta)$ imply this is a valid test.
- **Remark**: As noted by Romano & Shaikh (2008) ...

  ... generally conservative, i.e., may severely over cover $\lambda(\theta)$.
- Computationally difficult when $\dim(\theta)$ large.
- Related work by Kaido, Molinari & Stoye (2016) ...

  ... adjust critical value in $\phi_n(\theta)$ to avoid over-coverage.

Subvectors: subsampling

- **Main Idea**: Reject H_λ for large values of profiled test statistic:

$$T_n^{\text{prof}}(\lambda) = \inf_{\theta \in \Theta_\lambda} T_n(\theta) ,$$

  where $T_n(\theta)$ is one of test statistics from before.
- Use subsampling to estimate distribution of $T_n^{\text{prof}}(\lambda)$.
- High-level conditions for validity given by Romano & Shaikh (2008).
- **Remark**: Less conservative than proj., but choice of b problematic.

Subvectors: Minimum Resampling

- See Bugni, Canay & Shi (2014).
- Also rejects for large values of $T_n^{\text{prof}}(\lambda)$.
- In order to describe critical value, useful to define

$$J_n(x, \Theta_\lambda, s(\cdot), \lambda, P) = P\left\{ \inf_{\theta \in \Theta_\lambda} T\left(\hat{D}_n^{-1}(\theta) Z_n(\theta) + \hat{D}_n^{-1}(\theta) s(\theta), \hat{\Omega}_n(\theta)\right) \leqslant x \right\} .$$

Note

$$J_n(x, \Theta_\lambda, \sqrt{n}\mu(\cdot, P), \lambda, P) = P\{T_n^{\text{prof}}(\lambda) \leqslant x\} .$$

- **Old Idea**: Replace $s(\cdot)$ with 0 or $\hat{s}_n^{\text{gms}}(\cdot)$.

  – does not lead to valid tests.

- Indeed, for $P \in \mathbf{P}$ and $\lambda \in \Lambda_0(P)$,

$$\sqrt{n}\mu(\theta, P) \text{ need not be } \leqslant 0 \text{ for } \theta \in \Theta_\lambda \ .$$

$\Longrightarrow$ neither 0 nor $\hat{s}_n^{\text{gms}}(\cdot)$ provide (asymp.) upper bounds on $\sqrt{n}\mu(\cdot, P)$.

Failure of Naive GMS

Example

- Let $\{W_i\}_{i=1}^n = \{(W_{1,i}, W_{2,i})\}_{i=1}^n$ be i.i.d. $P = N(\mathbf{0}_2, I_2)$.
- Let $(\theta_1, \theta_2) \in \Theta = [-1,1]^2$ and consider

$$\mu_1(\theta, P) = E_P[\theta_1 + \theta_2 - W_{1,i}] \leqslant 0$$
$$\mu_2(\theta, P) = E_P[W_{2,i} - \theta_1 - \theta_2] \leqslant 0 \ .$$

- In this example

$$\Theta_0(P) = \{\theta \in \Theta : \theta_1 + \theta_2 = 0\} \ .$$

- Interested in testing the hypotheses

$$H_0 : \theta_1 = 0 \text{ vs. } H_1 : \theta_1 \neq 0 \ ,$$

  which corresponds to choosing $\lambda(\theta) = \theta_1$.
- In this case,

$$\Theta_\lambda = \{\theta \in \Theta : \theta_1 = 0, \theta_2 \in [-1,1]\} \ .$$

Failure of Naive GMS

Example (Test statistic)

- The profiled test statistic $T_n(\lambda)$ takes the form

$$T_n = \inf_{\theta_2 \in [-1,1]} T_n(0, \theta_2) = \inf_{\theta_2 \in [-1,1]} \left\{ \left[\frac{\theta_2 - \bar{W}_{n,1}}{\hat{\sigma}_{n,1}} \right]_+^2 + \left[\frac{\bar{W}_{n,2} - \theta_2}{\hat{\sigma}_{n,2}} \right]_+^2 \right\} ,$$

Note: $\hat{\sigma}_{n,j}(\theta)$ does not depend on θ in this example.

- Simple algebra shows

$$\theta_2^\star = \frac{\hat{\sigma}_{n,2}^2 \bar{W}_{n,1} + \hat{\sigma}_{n,1}^2 \bar{W}_{n,2}}{\hat{\sigma}_{n,2}^2 + \hat{\sigma}_{n,1}^2} \quad \text{w.p.a.1} ,$$

and this leads to

$$T_n = T_n(0, \theta_2^\star) = \frac{1}{\hat{\sigma}_{n,2}^2 + \hat{\sigma}_{n,1}^2} \left[\sqrt{n} \bar{W}_{n,2} - \sqrt{n} \bar{W}_{n,1} \right]_+^2 \xrightarrow{d} \frac{1}{2} [Z_2 - Z_1]_+^2 ,$$

where $(Z_1, Z_2) \sim N(\mathbf{0}_2, I_2)$. Both moments are asymp. binding and correlated.

Failure of Naive GMS

Example (Naive GMS)

- The naïve GMS approximation takes the form

$$T_n^{\text{naive}} = \inf_{\theta_2 \in [-1,1]} \left[-Z^*_{n,1} + s_1(0,\theta_2) \right]_+^2 + \left[Z^*_{n,2} + s_2(0,\theta_2) \right]_+^2 \ ,$$

where $\left\{ Z^*_{n,1}, Z^*_{n,2} | \{W_i\}_{i=1}^n \right\} \xrightarrow{d} Z = (Z_1, Z_2) \sim N(\mathbf{0}_2, I_2)$ w.p.a.1 .

- Some algebra shows that $\{T_n^{\text{naive}} | \{W_i\}_{i=1}^n\} \xrightarrow{d} \min\{[-Z_1]_+^2, [Z_2]_+^2\}$ w.p.a.1 . Follows from the fact that the GMS selection functions depend on

$$\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,1}^{-1}\bar{m}_{n,1}(0,\theta_2) = \kappa_n^{-1}\sqrt{n}\frac{\theta_2}{\hat{\sigma}_{n,1}} - \kappa_n^{-1}\sqrt{n}\frac{\bar{W}_{n,1}}{\hat{\sigma}_{n,1}}$$

$$\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,2}^{-1}\bar{m}_{n,2}(0,\theta_2) = \kappa_n^{-1}\sqrt{n}\frac{\bar{W}_{n,2}}{\hat{\sigma}_{n,2}} - \kappa_n^{-1}\sqrt{n}\frac{\theta_2}{\hat{\sigma}_{n,2}} \ .$$

- Naïve GMS: doesn't penalize large (+) values of $\kappa_n^{-1}\sqrt{n}\hat{\sigma}_{n,j}^{-1}(\theta)\bar{m}_{n,j}(\theta)$ (as $s_j(\theta) \leqslant 0$). It can afford to treat an ineq. as slack by making the other ineq. very positive (treat it as binding).

Subvectors: two ideas leading to one

- Lesson: naive GMS fails because $s \leqslant 0$ penalizes only one direction.
- Again: for $P \in \mathbf{P}$ and $\lambda \in \Lambda_0(P)$,

$$\sqrt{n}\mu(\theta, P) \text{ need not be } \leqslant 0 \text{ for } \theta \in \Theta_\lambda \ .$$

$\Longrightarrow$ neither 0 nor $\hat{s}_n^{\text{gms}}(\cdot)$ provide (asymp.) upper bounds on $\sqrt{n}\mu(\cdot, P)$.

- **Main Idea**:

  (a) Replace Θ_λ with a subset, e.g.,

$$\hat{\Theta}_n \approx \text{ minimizers of } T_n(\theta) \text{ over } \theta \in \Theta_\lambda \ ,$$

  over which $\hat{s}_n^{\text{gms}}(\cdot)$ provides asymp. upper bound on $\sqrt{n}\mu(\cdot, P)$.

  (b) Replace $s(\theta)$ with $\hat{s}_n^{\text{bcs}}(\theta) = (\hat{s}_{n,1}^{\text{bcs}}(\theta), \ldots, \hat{s}_{n,k}^{\text{bcs}}(\theta))'$ with

$$\hat{s}_{n,j}^{\text{bcs}}(\theta) = \frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\kappa_n \hat{\sigma}_{n,j}(\theta)} \ ,$$

  which does provide asymp. upper bound on $\sqrt{n}\mu(\cdot, P)$.

- Critical values from (a) and (b) both lead to valid tests.
- Combination of two ideas leads to even better test!

Minimum Resampling Test

Subvectors: Minimum Resampling

Let

$$T_n^{DR}(\lambda) \equiv \inf_{\theta \in \hat{\Theta}_n} S(Z_n^*(\theta) + s_n^{\text{gms}}(\theta), \hat{\Omega}_n(\theta)) ,$$

$$T_n^{PR}(\lambda) \equiv \inf_{\theta \in \Theta(\lambda)} S(Z_n^*(\theta) + s_n^{\text{bcs}}(\theta), \hat{\Omega}_n(\theta)) .$$

Let the critical value $\hat{c}_n^{MR}(\lambda, 1-\alpha)$ be the (conditional) $1-\alpha$ quantile of

$$T_n^{MR}(\lambda) \equiv \min \left\{ T_n^{DR}(\lambda), T_n^{PR}(\lambda) \right\} .$$

The minimum resampling test (or Test MR) is

$$\phi_n^{MR}(\lambda) \equiv 1 \left\{ T_n(\lambda) > \hat{c}_n^{MR}(\lambda, 1-\alpha) \right\} .$$

Remark: By combining both (a) and (b):

– Power advantages over both projection and subsampling

– Not true for (a) or (b) alone.

Main drawback is choice of κ_n.

Outline of Lecture

- Examples leading to moment inequalities
  - – Entry Games
  - – Revealed Preferences in Discrete Choice
  - – Missing data
- Confidence regions for partially identified models
  - – Importance of uniform asymptotic validity
- Moment inequalities: five distinct approaches
  1. Least Favorable Test
  2. subsampling
  3. Moment Selection
  4. Refined Moment Selection
  5. Two-step methods
- Subvector inference for moment inequalities
- Extensions

Many Moment Inequalities

- In many applications k may be large: motivates asymp. frameworks with $k = k_n \to \infty$
- Requires asymptotic approximations for normalized sums with increasing dimensions
- Recently developed by [5].
- Consider inference in models where $k_n \propto \exp(n^{\delta})$ for some $\delta > 0$.
- One-step tests: involve a "max"-type test statistic

$$\tilde{T}_n^{\max}(\theta) = \max_{1 \leqslant j \leqslant k} \frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} ,$$

and the following critical value

$$\hat{c}_{n,k}^{cck}(1-\alpha+2\beta,\theta) = \frac{\Phi^{-1}(1-(\alpha-2\beta)/k)}{\sqrt{1-\Phi^{-1}(1-(\alpha-2\beta)/k)^2/n}} .$$

Many Moment Inequalities

- Possible to improve on

$$\hat{c}^{cck}_{n,k}(1-\alpha,\theta) = \frac{\Phi^{-1}(1-\alpha/k)}{\sqrt{1-\Phi^{-1}(1-\alpha/k)^2/n}} .$$

  by incorporating information about $\sqrt{n}\mu(\theta, P)$.

- **Two-step tests:** uses a preliminary "selection" step.

  Step 1: the number of binding moments is estimated to be

$$\hat{k}_n = \sum_{j=1}^{k} \hat{s}^{\text{cck}}_{n,j}(\theta) ,$$

  where

$$\hat{s}^{\text{cck}}_{n,j}(\theta) = I\left\{ \frac{\sqrt{n}\bar{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)} > -2\hat{c}^{cck}_{n,k}(1-\beta,\theta) \right\}$$

  and $0 < \beta < \frac{\alpha}{3}$.

  Step 2: $\tilde{T}^{\max}_n(\theta)$ is compared with $\hat{c}^{cck}_{n,\hat{k}_n}(1-\alpha,\theta)$.

Conditional Moment Inequalities

- Many applications where the id. set involves **conditional** moment inequalities,

$$\Theta_0(P) = \{\theta \in \Theta : E_P[m(W_i,\theta)|Z_i] \leqslant 0 \ P\text{-}a.s.\} \ .$$

– see Andrews and Shi (2013) and Chernozhukov, Lee, and Rosen (2013) .

- **AS**: transform the cond. mom. ineq. into an infinite # of uncond. mom. ineq.
  - Can be done by choosing a set of weighting functions $\mathcal{G}$ with the property that

$$\Theta_{0,\mathcal{G}}(P) = \{\theta \in \Theta : E_P[m(W_i,\theta)g(Z_i)] \leqslant 0 \text{ for all } g \in \mathcal{G}\}$$

    is equal to $\Theta_0(P)$.

- **CLR**: "intersection bounds" interpretation. Let $\mathcal{V} \equiv \{(z,j) : z \in \mathcal{Z}, 1 \leqslant j \leqslant k\}$, and

$$\tilde{\mu}(\theta,P,v) = E_P[m_j(W_i,\theta)|Z_i = z] \ .$$

- Using this notation, the null hypotheses can be written as

$$H_\theta : \sup_{v \in \mathcal{V}} \tilde{\mu}(\theta,P,v) \leqslant 0 \ .$$

- CLR proposed a test based on non-parametric estimators of $E_P[m_j(W_i,\theta)|Z_i = z]$.

Random Set Theory

- Mathematical framework to study random objects whose realizations are **sets**.
- Useful for identification and inference when the object of interest is the identified set $\Theta_0(P)$.
- Well-developed area of mathematics. First application to partially identified models appeared in Beresteanu and Molinari (2006).
- Method useful when $\Theta_0(P)$ is a compact and convex set that is the Aumann expectation of a set-valued random variable.
- **Hypothesis**: For a given compact and convex set Ψ, the main inferential problem considered in the paper is testing

$$H_0 : \Theta_0(P) = \Psi$$

  versus the unrestricted alternative.
- **Test**: reject for large values of the normalized Hausdorff distance between Ψ and a sample analog of $\Theta_0(P)$ using a bootstrap critical value.
- Random set theory is particularly useful in providing tractable characterizations of (sharp) identified sets in partially identified models.

Bayesian Approach

- Identified models: frequentist confidence sets and Bayesian credible sets often **coincide**
- Equivalence **breaks down** in the context of partially identified models
- Priors on θ "influence" posterior inference statements concerning θ.
- Credible sets for θ thus tend to be **smaller** than frequentist CS.

  Result: from the Bayesian perspective, frequentist confidence sets are too wide, while from the frequentist perspective, Bayesian credible sets are too narrow.
- The lack of asymptotic harmony between Bayesian and frequentist inference is less severe when the object of interest is the identified set $\Theta_0(P)$ rather than $\theta \in \Theta_0(P)$.
- Recent papers propose robust Bayesian methods and show that a credible sets are also a valid frequentist confidence set for $\Theta_0(P)$.
- However: All results on "equivalence" are about "pointwise" asymptotic validity (a concern).

References

[1] Donald W. K. Andrews and P. Jia Barwick. Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica*, 80(6):2805–2826, November 2012.

[2] Donald W. K. Andrews and Xiaoxia Shi. Inference based on conditional moment inequalities. *Econometrica*, 81(2):609–666, 2013.

[3] Arie Beresteanu and Francesca Molinari. Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814, July 2008.

[4] F.A. Bugni, Ivan A. Canay, and X. Shi. Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*, 8(1):1–38, 2017.

[5] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Testing many moment inequalities. *arXiv preprint arXiv:1312.7614*, 2013.

[6] Victor Chernozhukov, Sokbae Lee, and Adam M Rosen. Intersection bounds: Estimation and inference. *Econometrica*, 81(2):667–737, 2013.

[7] Hyungsik R. Moon and Frank Schorfheide. Bayesian and frequentist inference in partially identified models. *Econometrica*, 80(2):755–782, 2012.

[8] Joseph P. Romano and Azeem M. Shaikh. Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference*, 138(9):2786–2807, September 2008.

[9] Joseph P. Romano, Azeem M. Shaikh, and Michael Wolf. A practical two-step method for testing moment inequalities. *Econometrica*, 82(5):1979–2002, 2014.

THE END!