Estimation and Inference in Mdoel with Partial Identification

from Chapter 5 in Handbook of Industrial Organization, Volume 4

Xinrui Liu

Hong Kong University of Science and Technology

March 4, 2025

Outline

Overview

A formal framework of identification

- **Model:** a model M is a set of functions or sets that satisfy some given restrictions.
  - e.g. restrictions on regression functions, distribution functions of errors or other unobservables, utility functions, payoff matrices, or information sets.
  - A model value $m \in M$ is an element of M. Each model value $m \in M$ implies a particular Data Generating Procss (DGP).
- **Data information:** ϕ is a set of constants and/or functions that we assume are known, or knowable, given the DGP.
  - Once we obtain the data, we immediately know ϕ.
  - e.g. data distribution function, or some features of distributions like conditional means, quantiles, autocovariances, or regression coefficients.
  - Let $\phi = \Pi(m)$ be a function (or a mapping) where $\Phi = \{\phi : \phi = \Pi(m), m \in M\}$ and $\Pi : M \to \Phi$.

Overview

- **Parameters of interest:** A set of parameters θ is a set of unknown constants and/or functions that characterize or summarize relevant features of a model.
  - Basically, θ is anything we might want to estimate.
  - e.g. regression coefficients, or the sign of an elasticity, or an average treatment effect.
  - Let $\theta = \Delta(m)$ be a function (or a mapping) where $\Theta = \{\theta : \theta = \Delta(m), m \in M\}$ and $\Delta : M \to \Theta$.
- **Structure:** The structure $s(\phi, \theta)$ is the set of all model values m that can yield both the given values ϕ and θ.
  - $s(\phi, \theta) = \{m \in M : \phi = \Pi(m), \theta = \Delta(m)\}$
  - $s(\phi, \theta)$ embodies the relationship between the parameters θ and what we could learn from data, which is ϕ.

Overview

Misspecification

The model M is defined to be misspecified if $\forall\ m \in M, \phi_0 \neq \Pi(m)$. Hence, the model M is not misspecified if $\exists\ m_0 \in M\ s.t.\ \phi_0 = \Pi(m_0)$.

Observationally Equivalent

θ and $\tilde{\theta}$ are said to be observationally equivalent in the model M if $\exists\ \phi \in \Phi \text{ s.t. } \mathrm{s}(\phi, \theta) \neq \varnothing\ and\ s(\phi, \tilde{\theta}) \neq \varnothing$. Equivalently, θ is observationally equivalent to $\tilde{\theta}$ if $\exists\ m, \tilde{m} \in M\ s.t.\ \theta = \Delta(m)\ and\ \tilde{\theta} = \Delta(\tilde{m})$.

- Let ϕ_0 be the value of ϕ that corresponds to the true DGP, that is, the DGP that actually generates what we can observe or know; m_0 doesn't conflict with what we can observe or know, which is ϕ_0; θ_0 is the true parameter of interest.

Overview

Identified Set

The identified set for θ_0 is $\Theta_I = \{\theta \in \Theta : \theta \textit{ is observationally equivalent to } \theta_0\}$.

- **Failure of Identification**
  - θ_0 is not identified if $\Theta_I = \emptyset$ i.e. the model is misspecified.
- **Point Identification**
  - θ_0 is point identified if $card(\Theta_I) = 1$ i.e. Θ_I is a singleton.
- **Partial Identification**
  - θ_0 is partially identified (or set identified) if $1 < card(\Theta_I) < card(\Theta)$ i.e. Θ_I is a proper subset of Θ.
- Other identifications: nonparametric identification, non-robust identification, nonstandard weak identification, sampling identification, semiparametric identification, structural identification, thin set identification ...
- Under partial identification, the identified set can have a complicate forms.

Overview

- Partial identification creates new and interesting issues for estimation and inference.
- **Questions:**
  - How do we estimate a set?
  - What is a "good" estimate of a set?
  - How do we construct a confidence region for a set?
  - Can we test an hypothesis about the true parameter under partial identification?
- For more complicated models where the identified set is difficult to describe explicitly, such questions are still the object of current research.
- We also discuss the difference between:
  - covering **a set** v.s. covering **any point** of the set
  - **pointwise** coverage v.s. **uniform** coverage

Overview

- We can try to obtain an estimate $\hat{\Theta}_I$ of the identified set Θ_I.
- Depending on the shape of the identified set, one can use different approaches to obtain such an estimate.
- Which **theoretical properties** should such an estimator $\hat{\Theta}_I$ have, independently of the method used to construct it?
- This issue needs clarification, as most standard notions from point estimation have no immediate counterpart for set estimation.

Overview

- At a minimum, such an estimator should be **consistent**.
- i.e. $\hat{\Theta}_I$ should get closer to Θ_I as the sample size increases:

  $$d(\hat{\Theta}_I, \Theta_I) \xrightarrow{p} 0$$

  for some distance measure $d(\cdot, \cdot)$ that works for sets.
- The literature on partial identification has most commonly used the **Hausdorff distance**.

Overview

Definition. (Hausdorff Distance)

The Hausdorff distance between sets A and B is

$$d_H(A,B) = \max\{\sup_{a\in A}\inf_{b\in B} d(a,b), \sup_{b\in B}\inf_{a\in A} d(a,b)\},$$

where $d(\cdot,\cdot)$ is a distance defined on the elements of the parameter space.

- The Hausdorff distance $d_H(A,B)$ is small when two conditions hold: every element $a \in A$ is "close" to at least one element of B and every element $b \in B$ is "close" to at least one element of A.
- Other representation:

$$\sup_{c\in A\cup B} |\inf_{a\in A} d(c,a) - \inf_{b\in B} d(c,b)|$$

.

Overview

Hausdorff Consistency

A sample analog estimator $\hat{\Theta}_I$ is consistent for Θ_I if

$$d_H(\hat{\Theta}_I, \Theta_I) \xrightarrow{p} 0 \ as \ n \to \infty.$$

- Consistency seems easy to achieve.
- Other common properties of point estimators, like **asymptotic normality** or **efficiency** are difficult to transfer to set estimation

Overview

Example (Interval Data)

This example is motivated by missing data problems, where Y is an unobserved real random variable bracketed below by Y_L and above by Y_U, both of which are observed real random variables. The parameter of interest $\theta = \mathbb{E}(Y)$ is known to satisfy the restriction

$$\mathbb{E}(Y_L) \leq \theta \leq \mathbb{E}(Y_U).$$

Hence the identified set is an interval, $\Theta_I = \{\theta \in \Theta : \mathbb{E}(Y_L) \leq \theta \leq \mathbb{E}(Y_U)\}$.

- **Question:** It is natural to use the **plug-in estimator** i.e. $\hat{\Theta}_I = [E_N(Y_L), E_N(Y_U)]$ where $E_N(\cdot)$ is the sample average. Is $\hat{\Theta}_I$ a consistent estimator for Θ_I?

Overview

- The Hausdorff distance here for the plug-in estimator is $d_H(\Theta_I, \hat{\Theta}_I) =$

$$\begin{cases} \max\{|E_N(Y_L) - \mathbb{E}(Y_L)|, |E_N(Y_U) - \mathbb{E}(Y_U)|\} & \textit{if } E_N(Y_L) \leq E_N(Y_U) \\ \textit{undefined} & \textit{if } E_N(Y_L) > E_N(Y_U) \end{cases}.$$

- **Case 1:** $\mathbb{E}(Y_L) < \mathbb{E}(Y_U)$
  Obviously, $P(E_N(Y_L) \leq E_N(Y_U)) \to 1$. By WLLN, we know $E_N(Y_L) \xrightarrow{p} \mathbb{E}(Y_L)$ and $E_N(Y_U) \xrightarrow{p} \mathbb{E}(Y_U)$. Then $d_H(\Theta_I, \hat{\Theta}_{I,N}) = max\{o_p(1), o_p(1)\} = o_p(1)$ which means $\hat{\Theta}_I$ is a consistent estimator.

Overview

- **Case 2:** $\mathbb{E}(Y_L) = \mathbb{E}(Y_U)$
  Now by CLT with with positive asymptotic variance, $P(E_N(Y_L) \leq E_N(Y_U)) \equiv P(\sqrt{N}(E_N(Y_L) - E_N(Y_U)) \leq 0) \approx \Phi(0) = \frac{1}{2}$. In this case, the event $\{d_H(\Theta_I, \hat{\Theta}_I)\ \textit{is undefined}\}$ has a non-vanishing probability, which means $d_H(\hat{\Theta}_I, \Theta_I) \overset{p}{\nrightarrow} 0$. Hence, $\hat{\Theta}_I$ is not consistent for the identified set.
- If Θ_I is a **non-degenerate** interval, then the plug-in estimator $\hat{\Theta}_I$ is consistent. However, if Θ_I is a singleton, equivalent to θ_0 being point identified, then $\hat{\Theta}_I$ is not consistent in general.
- The plug-in estimator may **perform poorly** in finite samples when $\mathbb{E}(Y_L) \approx \mathbb{E}(Y_U)$.
- **How to improve our estimation?**

Overview

- Consider the ϵ_n-**expansion** of the original plug-in estimator.
  - i.e. $\hat{\Theta}_{I,\epsilon} = [E_N(Y_L) - \epsilon_n, E_N(Y_U) + \epsilon_n]$ where ϵ_n is a sequence of positive numbers that converges to zero as $n \to \infty$.
- Now the distance for the modified estimator is $d_H(\Theta_I, \hat{\Theta}_{I,\epsilon_n}) =$

$$\begin{cases} |E_N(Y_L) - \mathbb{E}(Y_L) - \epsilon_n| \vee |E_N(Y_U) - \mathbb{E}(Y_U) + \epsilon_n| & \textit{if } E_N(Y_L) \leq E_N(Y_U) + 2\epsilon_n \\ \textit{undefined} & \textit{if } E_N(Y_L) > E_N(Y_U) + 2\epsilon_n \end{cases}$$

- **Claim:** If $\sqrt{n}\epsilon_n \to \infty$, then $\hat{\Theta}_{I,\epsilon_n}$ is a consistent estimator for the identified set.

Overview

- **Case 1:** $\mathbb{E}(Y_L) < \mathbb{E}(Y_U)$
  Obviously, $P(E_N(Y_L) - E_N(Y_U) \leq 2\epsilon_n) \geq P(E_N(Y_L) - E_N(Y_U) \leq 0) \rightarrow 1$. And if $E_N(Y_L) \leq E_N(Y_U)$, then $d_H(\Theta_I, \hat{\Theta}_{I,\epsilon_n}) = max\{o_p(1), o_p(1)\} = o_p(1)$ which means $\hat{\Theta}_{I,\epsilon_n}$ is a consistent estimator.
- **Case 2:** $\mathbb{E}(Y_L) = \mathbb{E}(Y_U)$
  Now by CLT with with positive asymptotic variance,
  $P(E_N(Y_L) - E_N(Y_U) \leq 2\epsilon_n) \equiv P(\sqrt{n}(E_N(Y_L) - E_N(Y_U)) \leq 2\sqrt{n}\epsilon_n) \approx \Phi(\infty) = 1$ for sufficiently large n. By the same argument, $\hat{\Theta}_{I,\epsilon_n}$ is a consistent estimator.
- That is to say, as long as the **rate of convergence** of ϵ_n is slower than $1/\sqrt{n}$, the ϵ_n-expansion of the plug-in estimator is consistent for the identified set.
- The formal framework will be discussed later.

Overview

- Given nominal size level α, we hope to construct a confidence set $CS_{n,\alpha}$.
- **Question:** What should the confidence set cover (asymptotically)?
- **The Identified Set v.s. Elements of the Identified Set**
  - Such a consideration does not arise in point identified models, when the identified set is a singleton.
  - Coverage of Θ_I: $P(\Theta_I \subseteq CS_{n,\alpha})$.
  - Coverage of $\theta \in \Theta_I$: $P(\theta \in CS_{n,\alpha})$.
- **Pointwise Inference v.s. Uniform Inference**
  - Pointwise means (implicitly) assuming a single fixed DGP.
  - Uniformity requires conditions that hold uniformly across a set of data generating processes i.e all $P \in \mathscr{P}$ where $\mathscr{P}$ is a space of possible DGPs.
  - Whether the sample size satisfying the coverage property depends on particular DGP or not.

Overview

- Different Targets: $CS_{n,\alpha}$ is asymptotically valid
  - **for pointwise coverage of the identified set**

$$\liminf_{n\to\infty} P(\Theta_I \subseteq CS_{n,\alpha}) \geq 1-\alpha.$$

  - **for pointwise coverage of the elements of the identified set**

$$\liminf_{n\to\infty} \inf_{\theta\in\Theta_I} P(\theta \in CS_{n,\alpha}) \geq 1-\alpha.$$

  - **for uniform coverage of the identified set**

$$\liminf_{n\to\infty} \inf_{P\in\mathscr{P}} P(\Theta_I \subseteq CS_{n,\alpha}) \geq 1-\alpha.$$

  - **for uniform coverage of the elements of the identified set**

$$\liminf_{n\to\infty} \inf_{P\in\mathscr{P}} \inf_{\theta\in\Theta_I} P(\theta \in CS_{n,\alpha}) \geq 1-\alpha.$$

- "$\geq$" means being **conservative** while "$=$" means being **not conservative**.

- By definition, a confidence set for the identified set is also a confidence set for the elements of the identified set, but a confidence set for the elements of the identified set is not necessarily a confidence set for the identified set.
- Imbens and Manski (2004) argue further that confidence regions for points in the identified set are generally of greater interest than confidence regions for the identified set itself, as there is still only one "true" value for θ in the identified set.
- Henry and Onatski (2012) provides a robust control argument for preferring inference for the identified set.

Methods: Moment Inequalities

- **Fact:** Some empirical models imply that the true value of the parameter θ_0 satisfies restrictions of the form $\mathbb{E}(m(W, \theta_0)) \geq 0$, where $m() = (m_1(\cdot), m_2(\cdot), \ldots, m_J(\cdot))$ is a vector-valued function that is known by the econometrician.
- **Moment Inequalities v.s. Moment Equalities**
  - Bisically, the moment inequality conditions are the generalization of the moment equality conditions.
  - e.g. $\mathbb{E}(m_j(W, \theta_0)) \geq 0$ and $\mathbb{E}(m_k(W, \theta_0)) \leq 0$ where $m_j(W, \theta_0) = m_k(W, \theta_0)$ implies $\mathbb{E}(m_j(W, \theta_0)) = \mathbb{E}(m_k(W, \theta_0)) = 0$ for some $j, k \in J$.
- **Unconditional Moment Inequalities v.s. Conditional Moment Inequalities**
  - For unconditional moment inequalities, $\Theta_I = \{\theta \in \Theta : \mathbb{E}(m(W, \theta)) \geq 0\}$.
  - For conditional moment inequalities, $\Theta_I = \{\theta \in \Theta : \mathbb{E}(m(W, \theta) | X = x) \geq 0 \textit{ for all } x\}$.

Methods: Criterion Function Approach

- **Fact:** Some empirical models can be characterized by a non-negative objective function $Q(\theta) \geq 0$ such that the identified set is $\Theta_I = \{\theta : Q(\theta) = 0\}$.
- The moment inequality approach is a **special case** of the criterion function approach.
- If θ_0 satisfies $\mathbb{E}(m(W, \theta_0)) \geq 0$, then we can take $Q(\theta) = \| \min(\mathbb{E}(m(W, \theta)), 0)\|^2$ such that $Q(\theta) = 0 \iff \mathbb{E}(m(W, \theta_0)) \geq 0$
- We can use sample objective function $Q_n(\theta)$ to estimates $Q(\theta)$: $Q_n(\theta) = \| \min(E_n(m(W, \theta)), 0)\|^2$.

Methods: Criterion Function Approach

- Estimation of a model based on moment inequality conditions is **straightforward**.
- Still be careful!
- First idea would be to estimate the identified set by $\tilde{\Theta}_I = \{\theta : Q_n(\theta) = 0\}$.
- But this does typically **not** work in applications!
- **Reason:** In finite samples, Q_n will often be positive with high probability even for values of θ within the identified set. Think of the interval data example!
- **Intuition:** Consider the standard GMM case with equalities and overidentification.
- $\tilde{\Theta}_I$ can be possibly empty when Θ_I is not, even in large samples.

Methods: Criterion Function Approach

- **A feasible approach: Chernozhukov et al.(2007)**
- Estimate Θ_I by the level set:

$$\hat{\Theta}_I = \{\theta : Q_n(\theta) \leq \epsilon_n\},$$

  where $\epsilon_n \to 0$ at an appropriate rate. (Recall the interval data example!)
- In most regular problems choosing $\epsilon_n = c\log(n)/n$ for some constant c is appropriate, and leads to an estimator of $\hat{\Theta}_I$ that is Hausdorff consistent.
- Under some technical conditions on Q_n, it can be shown that

$$d_H(\hat{\Theta}_I, \Theta_I) = O_p\left(\sqrt{\log(n)/n}\right).$$

- This is close to the $\sqrt{n}$ rate we typically get for parametric estimation problems under point identification.

Methods: Criterion Function Approach

- Inference is based on the contour set:

$$\mathcal{C}_{n,\alpha} = \{\theta : Q_n(\theta) \leq c_{n,\alpha}\},$$

  where $c_{n,\alpha}$ is a critical value (not a sequence).
- The critical value $c_{n,\alpha}$ can be chosen either by subsampling or based on asymptotic approximations for particular forms of the objective function.
- Chernozhukov et al.(2007) has shown that the contour set $\mathcal{C}_{n,\alpha}$ is asymptotically valid for the elements of the identified set:

$$\liminf_{n\to\infty} P(\theta' \in \mathcal{C}_{N,\alpha}) \geq 1 - \alpha \textit{ for any } \theta' \in \Theta_I$$

Methods: Criterion Function Approach

- In words, the expansion-based estimator has a nice asymptotic property.
- However, in many cases such an estimator is **biased** in **small** samples towards finding a too-small identified set, so it may be desirable to **bias-correct** the sampleanalog estimator.
- Bias-correction estimator: see Haile and Tamer (2003), Kreider and Pepper (2007), Andrews and Shi (2013), and Chernozhukov et al. (2013).
- Kaido and Santos (2014) study efficiency bounds when the moment inequalities and thus identified set is convex, finds the plug-in estimator is consistent, and proposes a valid bootstrap.

Methods: Moment Inequalities

- Inference of a model based on moment inequality conditions is **not straightforward**.
- **Main Challenge:** The asymptotic distributions of the proposed test statistics tend to be not pivotal i.e. they depend on unknown features of the DGP.
- Furthermore, the asymptotic distributions of the proposed test statistics depend on the "hold as equality" part of moment inequality conditions.
- **Possible Responses:** We are going to quickly introduce the three methods:
  - Least favorable approach(LF); from Rosen (2008).
  - Generalized moment selection approach (GMS); from Andrews and Soares (2010).
  - subsampling approach (SS); from Politis and Romano (1994).

Methods: Moment Inequalities

General Setup:

- The true value θ_0 satisfies

$$\mathbb{E}(m_j(Z,\theta)) \geq 0 \text{ for } j = 1, \ldots, p$$

$$\mathbb{E}(m_j(Z,\theta)) = 0 \text{ for } j = p+1, \ldots, p+v$$

  where $m(\cdot,\theta) = (m_j(\cdot,\theta), j = 1, \ldots, k)$ are known real-valued moment functions.
- θ_0 may or may not be identified by the moment conditions.
- Confidence sets for θ_0 can be constructed by inverting a test $T_n(\theta)$ for testing $H_0 : \theta = \theta_0$:

$$CS_n = \{\theta \in \Theta : T_n(\theta) \leq c(1-\alpha,\theta)\}$$

- Consider the sample moment functions $\bar{m}_n(\theta) = (\bar{m}_{n,1}(\theta), \ldots, \bar{m}_{n,k}(\theta))'$, where

$$\bar{m}_{n,j}(\theta) = \frac{1}{n}\sum_{i=1}^{n} m_j(Z_i,\theta) \text{ for } j = 1, \ldots, k$$

Methods: Moment Inequalities

- Let $\hat{\Sigma}(\theta)$ be an estimator of the asymptotic variance, $\Sigma(\theta)$, of $\sqrt{n}\bar{m}_n(\theta)$.
- For i.i.d. data we can take

$$\hat{\Sigma}(\theta) = \frac{1}{n}\sum_{i=1}^{n}(m(Z_i,\theta) - \bar{m}_n(\theta))(m(Z_i,\theta) - \bar{m}_n(\theta))'.$$

- For some S real function on $\mathbb{R}^p_{[+\infty]} \times \mathbb{R}^v \times \mathcal{V}_{k\times k}$ the statistic $T_n(\theta)$ is of the form

$$T_n(\theta) = S(n^{1/2}\bar{m}_n(\theta), \hat{\Sigma}(\theta)).$$

- $\mathbb{R}^p_{[+\infty]}$ is space of p-vectors whose elements are either real or $+\infty$.
- $\mathcal{V}_{k\times k}$ is the space of $k \times k$ matrices.

Methods: Moment Inequalities

- Different testing functions can be combined with different approaches to construct critical values.
- **General idea:** Under mild conditions, we have that

$$T_n(\theta) \xrightarrow{d} S(\Omega^{1/2} Z + h_1, \Omega).$$

- $Z \sim N(0_k, I_k)$ is a standard normal vector.
- $\Omega = \Omega(\theta)$ is the correlation matrix of $m(Z, \theta)$.
- h_1 is a k-vector with $h_{1,j} = 0$ for $j > p$ and $h_{1,j} \in [0, \infty]$ for $j \leq p$.
- Ideally, ideally one would use the $(1 - \alpha)$ quantile of $S(\Omega^{1/2} Z + h_1, \Omega)$.
- This requires knowledge of h_1, which cannot be estimated consistently.
- Different critical values are thus based on different approximations of $c_{h_1}(1 - \alpha, \theta)$.

Methods: Moment Inequalities

Approach 1: Least Favorable (LF)

- Rosen (2008) shows that distribution of $S(\Omega^{1/2}Z + h_1, \Omega)$ is stochastically largest when all moment inequalities are binding (i.e. hold as equalities).
- The **"worst case"** is thus that $h_1 = 0_k$, and the least favorable critical value is given by the $(1-\alpha)$ quantile of $S(\Omega^{1/2}Z + h_1, \Omega)$, denoted by $c_0(1-\alpha, \theta)$.
- With $\hat{D}_n(\theta) = \mathrm{diag}(\hat{\Sigma}_n(\theta))$ define $\hat{\Omega}_n(\theta) = \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}$.
- Then LF critical value is

$$c_{LF}(1-\alpha, \hat{\Omega}_n(\theta)) = \inf\{x \in \mathbb{R} : P(S(\hat{\Omega}_n(\theta)^{1/2}Z, \hat{\Omega}_n(\theta)) \leq x) \geq 1-\alpha\}$$

  for some random vector $Z \sim N(0_k, I_k)$ independent of the data.

Methods: Moment Inequalities

Results for LF Approach

- LF critical values are easy to implement, since they are very easy to compute.
- LF confidence sets are asymptotically valid in a uniform sense i.e.

$$\liminf_{n\to\infty} \inf_{P\in\mathscr{P}} P(\theta_0 \in CS_n^{LF}) \geq 1-\alpha$$

- LF critical values are conservative, since they are based on the least favorable case.
- LF critical values does not require determining a different critical value for each value of the parameter.

Methods: Moment Inequalities

Approach 2: Generalized Moment Selection (GMS)

- **Basic Idea:** To figure out which moment inequalities are binding from the data.
- For some $\kappa_n \to \infty$ at a suitable rate e.g. $\kappa_n = \sqrt{2\log(\log(n))}$, define

$$\xi_n(\theta) = \kappa_n^{-1} \hat{D}_n^{-1/2}(\theta) n^{1/2} \bar{m}_n(\theta).$$

- $\xi_n(\theta)$ is vector of normalized sample moments
- If $\xi_{n,j}(\theta)$ is **"large and positive"** then jth inequality "seems" not to be binding.
- If $\xi_{n,j}(\theta)$ is **"close to zero or negative"** then jth inequality "seems" to be binding.

Methods: Moment Inequalities

- GMS replaces h1 in limiting distribution by $\varphi(\xi_n(\theta), \hat{\Omega}_n(\theta))$.
- Function $\varphi = (\varphi_1, \ldots, \varphi_p, 0_v)$ can be chosen by the researcher, e.g.

$$\varphi_j^{(1)}(\xi, \Omega) = \infty \mathbb{I}\{\xi_j > 1\} \text{ (with } 0\infty = 0)$$
$$\varphi_j^{(2)}(\xi, \Omega) = (\xi_j)_+$$
$$\varphi_j^{(3)}(\xi, \Omega) = \xi_j$$

- GMS critical value is

$$c_{GMS}(1-\alpha, \hat{\Omega}_n(\theta), \kappa_n)$$
$$= \inf\{x \in \mathbb{R} : \Pr(S(\hat{\Omega}_n(\theta)^{1/2} Z + \varphi(\xi_n(\theta), \hat{\Omega}_n(\theta)), \hat{\Omega}_n(\theta)) \leq x) \geq 1-\alpha\}$$

for some random vector $Z \sim N(0_k, I_k)$ independent of the data.

Methods: Moment Inequalities

Results for GMS Approach

- GMS critical values are also easy to implement.
- GMS confidence sets are asymptotically valid in a uniform sense.

$$\liminf_{n\to\infty} \inf_{P\in\mathscr{P}} P(\theta_0 \in CS_n^{GMS}) \geq 1-\alpha.$$

- Under certain technical conditions, CS_n^{GMS} are not asymptotically conservative.
- CS_n^{GMS} have smaller volume than those based on CS_n^{LF}.
- CS_n^{GMS} depend on (arbitrary) choice of function φ.

Methods: Moment Inequalities

Approach 3: Subsampling (SS).

- Subsampling tries to approximate the distribution of $T_n(\theta)$ directly.
- **Basic Idea:** Suppose we could restart the data generating process as often as we wanted, and generated arbitrary many data sets $\{Z_i, i = 1, \ldots, n\}$. We could compute $T_n(\theta)$ for each new data set, and thus determine its distribution exactly.
- The infeasible approach version:
  - Draw small subsamples of size $b \ll n$ from the full data set (without replacement).
  - Compute test statistic for each subsample.
  - Use empirical distribution of subsample test statistics as an approximation to the distribution of $T_n(\theta)$.

Methods: Moment Inequalities

- Let b_n denote subsample size, which satisfies $b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$.
- There are $q_n = n!/((n-b_n)!b_n!)$ subsamples of size b_n.
- Let $T_{n,b,s}(\theta)$ be the test statistic on the sth subsample of size b_n.
- The empirical CDF of $T_{n,b,s}(\theta)$ is given by

$$U_{n,b}(x,\theta) = \frac{1}{q_n}\sum_{s=1}^{q_n} \mathbb{I}\{T_{n,b,s}(\theta) \leq x\}.$$

- SS critical value is

$$c_{SS}(1-\alpha,\theta,b) = \inf\{x \in \mathbb{R} : U_{n,b}(x,\theta) \geq 1-\alpha\}$$

Methods: Moment Inequalities

Results for SS Approach

- Computationally intensive, but works in theory under very weak conditions.
- SS confidence sets are asymptotically valid in a uniform sense and can be not asymptotically conservative under some certain conditions.
- SS test has less power than GMS test against certain local alternatives (and hence leads to asymptotically larger confidence sets).
- SS approximation can be unreliable in small or mid-size data sets.

Methods: Moment Inequalities

- There is a large literature on the advantages and disadvantages of different approaches to compute test statistics and critical values.
- Andrews and Jia (2011) recommend using a slightly modified version of the QLR statistic together with a particular GMS critical value.
- Bugni et al. (2011) study the properties of the confidence sets under local misspecification, finding that
  - LF critical values are more robust than GMS and SS critical values,
  - GMS and SS critical values are equally robust.
- There thus seems to be a **tradeoff between efficiency and robustness**.

Methods: Random Sets

- **Fact:** Many models exhibiting partial identification can be represented as involving a set of random variables that are compatible with the assumptions and the observed data.
- Sets are central to partial identification.
- Essentially, a set of random variables is a random set. Therefore, the random set framework is a natural generalization of the random variable framework.
- `Theory of Random Sets` (Molchanov, 2017)

Methods: Random Sets

Definition. (Random closed set)

A map X from a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to the family $\mathcal{F}$ of closed subsets of $\mathbb{R}^d$ is called a random closed set if

$$\{\omega \in \Omega : X(\omega) \cap K \neq \emptyset\} \in \mathscr{F}$$

where $\mathscr{F}$ is the σ-algebra on Ω for each compact set K in $\mathbb{R}^d$.

Definition. (Capacity functional)

A functional $T_X(K) : \mathcal{K} \mapsto [0, 1]$ given by

$$T_X(K) = \mathbb{P}\{X \cap K \neq \emptyset\}, \quad K \in \mathcal{K},$$

is called capacity functional (or hitting probability) of X.

Methods: Random Sets

Definition. (Measurable selection)

For any random set X, a (measurable) selection of X is a random element x with values in $\mathbb{R}^d$ such that $x(\omega) \in X(\omega)$ almost surely.

- Informally, the random set contains the random selections.
- A “selection” of a random set X is a random quantity x, also a mapping from the elements of the underlying probability space, whose realization is contained in X for every realization of the underlying probability space.

Methods: Random Sets

Theorem. (Artstein's inequality)

A probability distribution F on $\mathbb{R}^d$ is the distribution of a selection of a random closed set X in $\mathbb{R}^d$ if and only if

$$F(K) \leq T_X(K) = \mathbb{P}\{X \cap K \neq \emptyset\}$$

for all compact sets $K \subseteq \mathbb{R}^d$.

- This characterizes the possible distributions of the selection associated with a random set. For each such F, it is possible to construct selections x with distribution F that belongs to X almost surely.
- It says that the set of distributions of the set of selections of a random set are exactly those distributions F such that $F(K) \leq \mathbb{P}\{X \cap K \neq \emptyset\}$ for all compact sets K.

Methods: Random Sets

Theorem. (Law of large numbers for integrably bounded random sets)

Let $X, X_1, X_2, \ldots$ be *i.i.d.* integrably bounded random compact sets. Define $S_n = X_1 + \ldots + X_n$. Then

$$d_H\left(\frac{S_n}{n}, \mathbb{E}(X)\right) \to 0 \quad a.s. \text{ as } n \to \infty.$$

- If X is almost surely non-empty and its norm $\|X\| = \sup\{\|x\| : x \in X\}$ is an integrable random variable, then X is said to be **integrably bounded** and all its selections are integrable.

Methods: Random Sets

Theorem. (Central limit theorem for random closed sets)

Let $X, X_1, X_2, \ldots$ be *i.i.d.* copies of a random closed set X in $\mathbb{R}^d$ such that $\mathbb{E}\|X\|^2 < \infty$, and let $S_n = X_1 + \ldots + X_n$. Then as $n \to \infty$,

$$\sqrt{n} d_H \left(\frac{S_n}{n}, \mathbb{E}X \right) \Rightarrow \|\zeta\|_\infty = \sup \left\{ |\zeta(u)| : u \in \mathbb{S}^{d-1} \right\}$$

where ζ is a centered Gaussian random field on $\mathbb{S}^{d-1}$ with covariance function $\mathbb{E}\zeta(u)\zeta(v) = \mathbb{E}\|X\|^2 u \cdot v$, $u, v \in \mathbb{S}^{d-1}$.

Impelemention

- In point identified models, we hope to find the solution to
  - $Q_N(\theta) = 0$, or
  - $\max_{\theta \in \Theta} Q_N(\theta)$.
- In partially identified models, we hope to find **all** the solutions to
  - $Q_N(\theta) = 0$,or
  - $Q_N(\theta) \leq \epsilon_N$, or
  - $\max_{\theta \in \Theta} Q_N(\theta)$

Impelemention

- Grid search:
  - For many candidates θ', we should check whether $Q_N(\theta') \approx 0$ or $Q_N(\theta') \approx \max Q_N(\theta)$
  - Slow (especially for inference)!
  - Use parallel programming to speed up.
- Represent the identified set in a less computationally costly way:
  - e.g. $\Theta_I = \{\theta : E(Y_L) \leq \theta \leq E(Y_U)\}$, almost no computational burden.
  - e.g. try to write it in terms of a linear programming problem rather than a non-linear programming problem.
  - $\max_{\theta \in \Theta} Q_N(\theta)$
- Abandon some moment conditions ...

Impelemention

Kline and Tamer (2016)

- Simulation-based Bayesian approach.
- Moment inequalities: $P(Y|X = x) \leq m(x;\theta)$, where data on (Y, X) are available, $m(\cdot)$ is known.
- The identified set: $\Theta_I = \{\theta : P(Y|X = x) \leq m(x;\theta) \quad \forall x \in \mathcal{S}_X\}$, where $\mathcal{S}_X$ is the support of X.
- Basic idea:
    - firstly, construct posterior distributions on the finite dimensional vector $P(Y|X)$
    - use draws from this posterior to construct a posterior distribution for Θ_I via the identified set mapping.

Impelemention

Chen et al. (2018)

- Monte Carlo simulation methods from a quasi-posterior.
- Start with an optimal objective function $L(\theta)$, e.g.
  - a likelihood, or
  - an optimally weighted GMM, or
  - an optimal moment inequality objective function.
- The identified set:
$$\Theta_I = \left\{ \theta \in \Theta : L(\theta) = \sup_{\vartheta \in \Theta} L(\vartheta) \right\}.$$
  It can be a singleton.
- $L_N(\theta)$ is a sample analog of $L(\theta)$. Let $\hat{\theta} \in \Theta$ denote an approximate maximizer of L_N i.e. $L_n(\hat{\theta}) = \sup_{\theta \in \Theta} L_n(\theta) + o_p(n^{-1})$.

Impelemention

- To construct confidence sets for the identified set:
  - Draw a sample $\{\theta^1, \ldots, \theta^B\}$ from the quasi-posterior distribution Π_n:

$$\mathrm{d}\Pi_n(\theta|\mathbf{X}_n) = \frac{e^{nL_n(\theta)}\mathrm{d}\Pi(\theta)}{\int_\Theta e^{nL_n(\theta)}\mathrm{d}\Pi(\theta)}.$$

  - Calculate the $(1-\alpha)$ quantile of $\{L_n(\theta^1), \ldots, L_n(\theta^B)\}$, denoted by $\zeta_{n,\alpha}^{\mathrm{mc}}$.
  - The $100\alpha\%$ confidence set for Θ_I is then

$$\hat{\Theta}_\alpha = \{\theta \in \Theta : L_n(\theta) \geq \zeta_{n,\alpha}^{\mathrm{mc}}\}.$$

Impelemention

What to report?

- If $\theta = (\theta_1, \theta_2, \dots, \theta_K)$ is the parameter of the model,
  - the identified set for θ is Θ_I.
  - the identified set for θ_k is $\Theta_{I,k}$.
- If $dim(\theta) \leq 3$, report the identified set graphically.
- Otherwise , partition θ into $\theta = (\theta^{(1)}, \theta^{(2)})$ such that $dim(\theta^{(1)}) \leq 3$. Then report $\theta^{(1)}$ graphically at various specified values for $\theta^{(2)}$.

Conclusion

- **The law of decreasing credibility**
  - The credibility of inference decreases with the strength of the assumptions maintained. ——Manski (2003)
  - the strength of assumptions v.s. credibility of the results.
  - Avoid taking things to the extreme: Partial identification results are not *per se* necessarily better than point identification results.
- **Misspecification in model with moment inequalities**
- **Limited software implementations**
  - Stata packages: `clrbound`, `tebounds`, `cmi_test` ...