Introduction to Partial Identification

C. Bontemps[1]

[1]Toulouse School of Economics

DEEQA - Winter 2014

Partial Identification in Econometrics I

- The traditional approach in statistics and econometrics is to consider models that are **point identified**.
- Given an infinite amount of data, in such models one can always infer without uncertainty what the values of the objects of interest are.
- Uncertainty about the true value of a parameter is thus only due to using a finite data set.
- Researchers traditionally felt uncomfortable about models in which point identification fails.
- They have therefore often added additional assumptions to their models that have identifying power, even if they are not well justified by economic theory.
- **Problem:** Empirical results might be driven by a priori assumptions, and not by the data.
  - Two researchers using the same data set might come to different conclusions, depending on which additional assumptions they impose.

Partial Identification in Econometrics II

- It is therefore important to find out what conclusions can be drawn about a research question under **weaker** or **minimal** assumptions.
- This sometimes means that one has to give up point identification.
- That is, one has to work with a model where it is not possible to infer the true value of the parameter of interest even with an infinitely large data set.
- **Such models are not useless!**
- The data might reveal some non-trivial insights about the objects of interest, even though they do not allow for an exact quantification.
- This perspective is called *partial identification*.

- Partial identification occurs in many areas of applied econometrics:
  - measurement error model,
  - missing data models,
  - treatment effects,
  - market entry games,
  - Economic models with inequalities.

Partial Identification in Econometrics III

- Partial identification analysis is about finding out which values of the true parameter of interest are compatible with the observations we made.
    - How can we obtain the identified set?
- Partial identification also poses new challenges for estimation and inference:
    - How can we obtain "estimates" in a setting where consistent estimation is impossible?
    - How can we test an hypothesis about the parameter of interest?

Linear Error-in-Variables Models I

- Frisch (1934) studied linear regression problems when variables are measured with error.
- Suppose that there is a linear model

$$Y^* = \beta_1 + \beta_2 X^* + \varepsilon$$

where Y^*, X^*, ε are scalar and $E[\varepsilon] = 0$, $E[X^* \varepsilon] = 0$.
- Assume that both Y^* and X^* are observed with error:

$$Y = Y^* + \Delta Y$$
$$X = X^* + \Delta X$$

- Here ΔY and ΔX are unobserved measurement errors that are uncorrelated with the other primitives of the model.
- **Question:** What can we learn from observing (Y, X) about the slope parameter β_2 in the true regression?

Linear Error-in-Variables Models II

Inconsistency of conventional regression

The true model implies

$$Y = \beta_1 + \beta_2 X + \underbrace{\varepsilon + \Delta Y - \beta_2 \Delta X}_{W}.$$

and if we regress Y on X:

$$\text{plim}\, \hat{\beta}_2 = \beta_2 + \frac{\mathrm{Cov}(X, W)}{\mathrm{Var}(X)}.$$

Moreover:

$$\begin{aligned} \mathrm{Cov}(X, W) &= \mathrm{Cov}(X^* + \Delta X, \varepsilon + \Delta Y - \beta_2 \Delta X) \\ &= -\beta_2 \mathrm{Var}(\Delta X). \end{aligned}$$

Linear Error-in-Variables Models III

so that

$$\text{plim}\,\hat{\beta}_2 = \beta_2 - \beta_2 \frac{\mathrm{Var}(\Delta X)}{\mathrm{Var}(\Delta X) + \mathrm{Var}(X^*)} = \beta_2 \frac{\mathrm{Var}(X^*)}{\mathrm{Var}(\Delta X) + \mathrm{Var}(X^*)}.$$

Linear Error-in-Variables Models IV

Observations

- Standard regression is inconsistent in the presence of measurement error.
- The slope coefficient is biased towards zero (i.e. any relationship is attenuated).
- Rejection of significance is still reliable but the power is reduced.
- Classical measurement error in the dependent variable has no effect
- Caution: Direction of the bias is not obvious in multivariate settings.

Linear Error-in-Variables Models V

One solution: bounds on β_2 Notice that there exists other solutions (outside information related to the error variance, instrumental variables that are not correlated with the measurement error, higher-order moments, repeated measurements,etc.). See the papers of S. Schennach and/or J. Hu.

The structure of the model gives three equations related to the second moments of observables:

$$\mathrm{Var}(Y) = \beta_2^2 \mathrm{Var}(X^*) + \mathrm{Var}(\Delta Y) \quad (1)$$

$$\mathrm{Var}(X) = \mathrm{Var}(X^*) + \mathrm{Var}(\Delta X) \quad (2)$$

$$\mathrm{Cov}(X, Y) = \beta_2 \mathrm{Var}(X^*) \quad (3)$$

$$(1) + (3) \rightarrow \ \mathrm{Var}(Y) = \beta_2 \mathrm{Cov}(X, Y) + \mathrm{Var}(\Delta Y) \quad (4)$$

Linear Error-in-Variables Models VI

Two inequalities can be derived using the fact that a variance is positive:

- $\mathrm{Var}(\Delta X) \geq 0$ and, in this case,

$$\beta_2 \geq \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}$$

.

- $\mathrm{Var}(\Delta Y) \geq 0$ and, in this case,
- The identified set of the slope coefficient is thus given by

$$\left[\frac{\mathrm{Cov}(Y, X)}{\mathrm{Var}(X)}, \frac{\mathrm{Var}(Y)}{\mathrm{Cov}(Y, X)}\right].$$

Linear Error-in-Variables Models VII

Comments

- This set is *sharp*: no value in this set, including the end points, can be rejected as the true slope parameter β_0.
- We get the upper and lower bound of $\mathrm{Var}(\Delta X) = 0$ and $\mathrm{Var}(\Delta Y) = 0$, respectively.
- Even with large samples, we cannot point identify the slope value.
- To obtain point identification, literature on measurement errors uses two principal approaches

Road Map

1. Some (usual) examples of interest
2. The Moment inequality approach
   - The original paper
   - Andrews and co.
   - further discussion
3. Convexity and the support function approach
4. Conclusion and perspective

A more formal definition I

- Consider an observable random vector $Z \in \mathbb{R}^d$, distributed according to some probability measure P_0, i.e. $F_Z(z) = P_0(Z \leq z)$.
- Let $\mathcal{P}$ be a *model* for the underlying probability measure. That is, we assume that $P_0 \in \mathcal{P}$.
- A model is simply the collection of all probability measures that are compatible with the assumptions we make.
- In addition to the observable random vector $Z \in \mathbb{R}^d$, there may also be unobservable random objects whose distribution is also determined by the probability measure P_0.
- Suppose we want to learn $\theta_0 = \Gamma(P_0)$.
- This parameter could be finite or infinite dimensional, taking values in the space $\Theta = \{\Gamma(P) : P \in \mathcal{P}\}$.

- **Point identification approach:** Show that θ_0 can be written in terms of the distribution of observed outcomes, i.e. $\theta_0 = \nu(F_Z)$.
- In partially identified models, such a relationship may not exist.

A more formal definition II

- There might be probability measures $P, P' \in \mathcal{P}$ such that $\Gamma(P) \neq \Gamma(P')$, but $P(Z \leq z) = P'(Z \leq z)$ for all values of z.
- In this case, we are unable to pin down the exact value of θ_0 even in large samples, but we might be able to learn the values that are compatible with the distribution of observables.
- These are given by the set

$$\Theta_I = \{\Gamma(P) : P \in \mathcal{P} \text{ and } P(Z \leq z) = F_Z(z) \text{ for all } z\}.$$

- We call Θ_I the *identified set*.
- Furthermore, we say that
  - θ_0 is *point identified* if Θ_I is a singleton,
  - θ_0 is *not identified* if $\Theta_I = \Theta$,
  - θ_0 is *partially identified* if $\Theta_I \subset \Theta$.
- Under partial identification, the identified set can have a complicate forms.

A more formal definition III

- One of the most important challenges when working with partially identified models is to find a simple characterization for the identified set.

Examples

- Partial identification issues were studied as early as in the 1930's (and probably earlier).
- This research had little impact on applied economics.
  - Recent interest in partial identification started with the work of Manski in the 1990's.
- We now consider two of the leading examples considered in this literature. Additional examples are
  - Bounds on the Joint CDF with given Marginals
  - Missing Data and Treatment Effects
  - Linear Models with Interval censored regressors

Example 1: Linear Models with Interval Data I

- Consider the model $Y = X'\theta_0 + \varepsilon$ when the outcome variable is **interval measured**.
  - We do not observe Y directly but (Y_l, Y_u) such that $P(Y \in [Y_l, Y_u]) = 1$.
  - We assume (for the sake of simplicity) uncorrelation between ε and X.
  - We need to assume that Y is bounded.
  - Otherwise we have all the usual assumptions for linear regression.
- Interval censoring is common in economic applications.
- Income, wealth, wages, hours of work, taxes, etc. are often only measured in brackets.
  - Often due to data confidentiality reasons.
  - Also increases response rate in surveys.
- **Object of interest** is the parameter θ_0.

Example 1: Linear Models with Interval Data II

We start with the assumption that X is univariate. We can easily characterize the identified set:

$\Theta_I = \{t \in \mathbb{R}$ such that there exists a r.v. $\lambda \in [0,1] E(X(Y_l + \lambda(Y_u - Y - l))) = E($

- The identified set is a closed interval centered in $E(\frac{Y_u+Y_l}{2})/E(X^2)$.
-

$$\Theta_I = E(X\frac{Y_u + Y_l}{2})/E(X^2) \pm E(\left|X\frac{Y_u - Y_l}{2}\right|)/E(X^2).$$

More generally

$\Theta_I = \{t \in \mathbb{R}$ such that $t = E(XX')^{-1}E(X(Y_l + \lambda(Y_u - Y - l)))$ for some r.v. in $[0$

Example 2: 2x2 Entry model I

Two firms, A and B, contest a set of markets.

- In market m, where $m = 1, \dots, M$, the profits for firms A and B are

$$\pi_{Am} = \alpha_A + \delta_A d_{Bm} + \varepsilon_{Am}$$
$$\pi_{Bm} = \alpha_B + \delta_B d_{Am} + \varepsilon_{Bm},$$

  where $d_{Fm} = 1$ if firm F is present in market m, for $F \in \{A, B\}$, and zero otherwise.
- A more realistic model would also include observed market and firm characteristics.
- Firms enter market m if their profits in that market are positive.

Example 2: 2x2 Entry model II

- Firms observe all components of profits, including those that are unobserved to the econometrician, $(\varepsilon_{Am}, \varepsilon_{Bm})$, and so their decisions satisfy:

$$d_{Am} = \mathbb{I}\{\pi_{Am} \geq 0\}$$
$$d_{Bm} = \mathbb{I}\{\pi_{Bm} \geq 0\}.$$

- The unobserved components of profits, ε_{Fm}, are independent across markets and firms.
- The econometrician observes in each market only the pair of indicators d_A and d_B.

- For simplicity, we assume that δ_A and δ_B are negative, and that $(\varepsilon_{Am}, \varepsilon_{Bm})$ has a distribution F_Ω that is known up to finite-dimensional parameter Ω.
- Our aim is to learn the vector of parameters $\theta = (\alpha_A, \alpha_B, \delta_A, \delta_B, \Omega)$.

Example 2: 2x2 Entry model III

- With distributional assumptions on $(\varepsilon_{Am}, \varepsilon_{Bm})$, it seems we could obtain parameters of interest by maximizing the likelihood function of the problem.
- That is, we could try to choose parameter θ such that we match the observed four choice probabilities $p_{ij} = P(d_A = i, d_b = j)$ as good as possible.
- But this is not the case: for pairs of $(\varepsilon_{Am}, \varepsilon_{Bm})$ such that

$$-\alpha_A \leq \varepsilon_{Am} \leq -\alpha_A - \delta_A$$
$$-\alpha_B \leq \varepsilon_{Bm} \leq -\alpha_B - \delta_B$$

  both $(d_A, d_B) = (0,1)$ and $(d_A, d_B) = (1,0)$ satisfy the profit maximization condition.
- Multiple equilibria are possible in this region.
- In the terminology of this literature, the model is not complete. (DRAW PICTURE)

Example 2: 2x2 Entry model IV

- **Consequence:** the probability of the outcome $(d_A, d_B) = (0,1)$ cannot be written as a function of the parameters of the model, $\theta = (\alpha_A, \alpha_B, \delta_A, \delta_B, \Omega)$, even given distributional assumptions on $(\varepsilon_{Am}, \varepsilon_{Bm})$.
- This would require an **equilibrium selection rule**.
- Instead the model implies a lower and upper bound on this probability:

$$H_L^{(0,1)}(\theta) \leq \Pr((d_A, d_B) = (0,1)) \leq H_U^{(0,1)}(\theta)$$

where

$$H_L^{(0,1)}(\theta) = \Pr(\varepsilon_{Am} < -\alpha_A, -\alpha_B < \varepsilon_{Bm}) \\ + \Pr(-\alpha_A < \varepsilon_{Am} < -\alpha_A - \delta_A, -\alpha_B - \delta_B < \varepsilon_{Bm})$$

and

$$H_U^{(0,1)}(\theta) = \Pr(\varepsilon_{Am} < -\alpha_A - \delta_A, -\alpha_B < \varepsilon_{Bm}).$$

Example 2: 2x2 Entry model V

- Similar bounds can then be obtained on the probability of the event that $(d_A, d_B) = (1,0)$.
- The probability that $(d_A, d_B) = (1,1)$ and $(d_A, d_B) = (0,0)$ can be exactly determined.

- The identified set is thus given by:

$$\begin{aligned}\Theta_I = \{\theta : \quad & H_L^{(0,1)}(\theta) \leq \Pr((d_A, d_B) = (0,1)) \leq H_U^{(0,1)}(\theta), \\ & H_L^{(1,0)}(\theta) \leq \Pr((d_A, d_B) = (1,0)) \leq H_U^{(1,0)}(\theta), \\ & \Pr((d_A, d_B) = (0,0)) = H^{(0,0)}(\theta), \\ & \Pr((d_A, d_B) = (1,1)) = H^{(1,1)}(\theta)\}\end{aligned}$$

- In general, this set does not have a more simple characterization.
- Beresteanu et al. (2011) and Galichon and Henry (2011) discuss alternative characterizations based on the theory of random sets.

Example 2: 2x2 Entry model VI

- Tamer (2003) shows that if profits are of the form

$$\pi_{Fm} = \alpha_F + X'_{Fm}\beta_F + \delta_F d_{Fm} + \varepsilon_{Fm}, \quad F \in \{A, B\},$$

where X_F are observable firm characteristics, one can achieve point identification under a large support condition on one of the characteristics ("identification at infinity").

General comments I

- Partial identification creates new and interesting issues for estimation and inference.
  - 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 or any point of the set
- Uniformity of the approach with respect to the (true but unknown) size of the set.

General comments II

- One can try to obtain an estimate $\widehat{\Theta}_I$ of the identified set Θ_I.
- Depending on the shape of the identified set, one can use different approaches to obtain such an estimate (we talk about this soon).
- **Question:** Which theoretical properties should such an estimator $\widehat{\Theta}_I$ 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.

General comments III

- At a minimum, such an estimator should be **consistent**.
- **What does this mean?**
- As the sample size increases, $\widehat{\Theta}_I$ should get closer to Θ_I:

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

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

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

- Other distance measures are possible in principle, but are rarely considered.
- Other common properties of point estimators, like **asymptotic normality** or **efficiency** are difficult to transfer to set estimation.

Estimation: Interval Identified Parameters I

- Estimation is straightforward if the true parameter θ_0 is scalar and the identified set takes the form of an interval, i.e.

$$\Theta_I = [\theta_l, \theta_u].$$

- In this case, we can estimate Θ_I by

$$\widehat{\Theta}_I = [\hat{\theta}_l, \hat{\theta}_u],$$

  where $\hat{\theta}_l, \hat{\theta}_u$ are suitable estimates of the upper and lower boundary of the interval.
- It is straightforward to show that if $(\hat{\theta}_l, \hat{\theta}_u) \xrightarrow{p} (\theta_l, \theta_u)$ the set estimator $\widehat{\Theta}_I$ is consistent in the Hausdorff norm.
- **Proof:** $d_H(\hat{\Theta}_I, \Theta_I) = max\{\|\hat{\theta}_l - \theta_l\|, \|\hat{\theta}_u - \theta_u\|\} \xrightarrow{p} 0$

Estimation: Interval Identified Parameters II

- **Example (Interval censoring with one explanatory variable):** Observe (Y_{ui}, Y_{li}, X_i), where $Y_i \in [Y_{li}, Y_{ui}]$ is the outcome of interest.
- Recall that the identified set is

$$\Theta_I = \frac{E(\frac{X(Y_u+Y_l)}{2})}{E(X^2)} \pm \frac{E(\left|X\frac{Y_u-Y_l}{2}\right|)}{E(X^2)}.$$

- **Estimation** by sample analogues: put

$$\hat{\theta}_c = \frac{\frac{1}{n}\sum_{i=1}^n X_i(Y_{ui}+Y_{li})}{2\frac{1}{n}\sum_{i=1}^n X_i^2} \text{ and } \hat{h}l = \frac{\frac{1}{n}\sum_{i=1}^n |X_i|(Y_{ui}-Y_{li})}{2\frac{1}{n}\sum_{i=1}^n X_i^2}$$

and set

$$\hat{\theta}_l = \hat{\theta}_c - \hat{h}l \text{ and } \hat{\theta}_u = \hat{\theta}_c + \hat{h}l.$$

- Consistency follows from Law of Large Numbers and Continuous Mapping Theorem.

Confidence region in a point identified model

In a point identified model, a confidence region of nominal size asymptotically equal to $1-\alpha$ can be derived from a test statistic ξ whose aims is to test:

$$H_0 : \theta = \theta_0 \text{ against } H_a : \theta \neq \theta_0.$$

Following Lehmann (1986, Chapter 3), the confidence region $CI^n_{1-\alpha}$ is the collection of parameters $\theta \in \mathbb{R}^d$ for which the null hypothesis is not rejected *i.e.*.

$$\lim_{n\to+\infty} Pr\left(\theta^0 \in CI^n_{1-\alpha}\right) = 1-\alpha.$$

Inference: A Simple Example I

- Consider a stylized setting where θ_0 is scalar and interval-identified:

$$\Theta_I = [\theta_l, \theta_u].$$

- The upper and lower bound can be estimated by $\hat{\theta}_u$ and $\hat{\theta}_l$, resp., which satisfy:

$$\sqrt{n}((\hat{\theta}_u, \hat{\theta}_l) - (\theta_u, \theta_l)) \xrightarrow{d} N(0, \Sigma)$$

where Σ is a diagonal matrix with $\mathrm{diag}(\Sigma) = (\sigma_u^2, \sigma_l^2)$.
- We thus have asymptotically normal and independent estimated of the upper and lower boundaries of the identified set.
- Let $q(\alpha)$ denote the α quantile of the standard normal distribution.
- By analogy to point identified case, we consider confidence sets of the form

$$CS_n(a) = [\hat{\theta}_l - q(a)\hat{\sigma}/\sqrt{n}, \hat{\theta}_u + q(a)\hat{\sigma}/\sqrt{n}] \tag{5}$$

for some $a > 1/2$.

Inference: A Simple Example II

- Choice of a depends in desired properties of confidence set.
- Choose $a = \sqrt{(1-\alpha)}$ for $CS_n(a)$ to be $(1-\alpha)$ confidence set for Θ_I.
- **Proof:** Just calculate the probability:

$$\Pr(\Theta_I \in CS_n(a)) = \Pr(\theta_l \geq \hat{\theta}_l - q(a)\hat{\sigma}_l/\sqrt{n} \text{ and } \theta_u \leq \hat{\theta}_u + q(a)\hat{\sigma}_u/\sqrt{n}) \to a$$

- Getting a confidence interval for θ_0 is slightly more complicated.
  - Suppose $\theta_l < \theta_0 < \theta_u$. Then $\Pr(\theta_0 \in CS_n(a)) \to 1$ for all values $a > 1/2$.
  - Suppose $\theta_l = \theta_0$. Then $\Pr(\theta_0 \in CS_n(a)) \to a$.
  - Suppose $\theta_u = \theta_0$. Then $\Pr(\theta_0 \in CS_n(a)) \to a$.
- We thus have that

$$\liminf_{n\to\infty} \inf_{\theta_0 \in \Theta_I} \Pr(\theta_0 \in CS_n(1-\alpha)) \geq 1-\alpha.$$

Inference: A Simple Example III

- $CS_n(1-\alpha)$ looks like a "good" $(1-\alpha)$ confidence set for θ_0.
- **Problem:** Suppose that $\theta_l = \theta_u$, and that $\hat{\theta}_l = \hat{\theta}_u$ in this case.
  - This is valid, because point identification is just a special case of partial identification.
- Then we find that

$$\Pr(\theta_0 \in CS_n(1-\alpha)) = \Pr(|\sqrt{n}(\hat{\theta}-\theta_0)/\hat{\sigma}| \leq q(1-\alpha)) \to 1-2\alpha$$

- The confidence interval is too liberal in this case.
- A similar argument applies when θ_l and θ_u are not equal, but close together.
- The confidence set would be **shorter** than under point identification in this case.
- **Reason:** $CS_n(1-\alpha)$ is not a valid $(1-\alpha)$ confidence set for θ_0 **uniformly** over the length $p = \theta_u - \theta_l$ of the identified set:

$$\liminf_{n\to\infty} \inf_{p} \inf_{\theta_0 \in \Theta_I} \Pr(\theta_0 \in CS_n(1-\alpha)) \geq 1-2\alpha.$$

Inference: A Simple Example IV

- $CS_n(1-\alpha/2)$ would be a valid $(1-\alpha)$ confidence set for θ_0 uniformly over p.
- But this would be very conservative when p is large.
- Imbens and Manski (2004) suggest to adjust the critical value based on an estimate $\hat{p} = \hat{\theta}_u - \hat{\theta}_l$ of the length of the identified set.
- Define $CS_n^{IM} = [\hat{\theta}_l - C_n\hat{\sigma}/\sqrt{n}, \hat{\theta}_u + C_n\hat{\sigma}/\sqrt{n}]$
- Here the critical value C_n satisfies

$$\Phi\left(C_n + \sqrt{n}\frac{\hat{p}}{\hat{\sigma}}\right) - \Phi(-C_n) = 1-\alpha.$$

  and $C_n = q(1-\alpha/2)$ if $\hat{p} = 0$.
- One can show that CS_n^{IM} has asymptotic coverage rate of $1-\alpha$, uniformly over p.
- Note that $C_n \in (q(1-\alpha), q(1-\alpha/2))$ for every value of $\hat{p}$.

Interval censoring case

$\Theta_I = \{t \in \mathbb{R}$ such that $t = \mathbb{E}(XX')^{-1}\mathbb{E}(X(Y_l + \lambda(Y_u - Y - l)))$ for some r.v. in [0

- Several approaches are possible.
- **Approach 1:** Criterion approach like the Modified Minimum Distance (Manski and Tamer, 2002).
- Monotonicity of the problem implies that

  $$\Theta_I = \{\theta \in \mathbb{R}^k : \mathbb{E}X(Y_l + Y_u)/2 - \mathbb{E}|X_j|(Y_u - Y_l)/2 \leq (\mathbb{E}(XX')\theta)_j \leq \mathbb{E}X(Y_l +$$

- Can write Θ_I as the argmin of an objective function:

  $$\Theta_I = \underset{\theta}{\text{argmin}} \sum_j (U_j - (\mathbb{E}(XX')\theta)_j)_-^2 + ((\mathbb{E}(XX')\theta)_j - L_j)_-^2))$$

  with $(a)_+ = a\mathbb{I}\{a \geq 0\}$ and $(a)_- = a\mathbb{I}\{a \leq 0\}$.
- **Approach 2:** Support functions (e.g. Bontemps et al., 2011). See also Beresteanu and Molinari (2008).
- One can show that the set Θ_I is bounded and convex.
- It can thus equivalently be described through its support function:

Estimation: Criterion Function Approach I

- When the shape of the identified set is more complicated, other techniques have to be used.
- Chernozhukov, Hong and Tamer (2007, Ecma) generalize the concept of extremum estimators to settings with partial identification.
- They study the case where

$$\Theta_I = \operatorname*{argmin}_{\theta \in \Theta} Q(\theta)$$

and there exists a well-defined sample objective function $Q_n(\cdot)$ such that

$$\sup_{\theta} \| Q_n(\theta) - Q(\theta) \| \xrightarrow{p} 0.$$

Estimation: Criterion Function Approach II

- Assume without loss of generality that $Q(\theta) \geq 0$ for all θ, and that $Q(\theta) = 0$ if $\theta \in \Theta_I$.
- Examples include Modified Minimum Distance approach of Manski and Tamer (2002), and more generally Moment Inequality Models.

Estimation: Criterion Function Approach III

- **Remark (Moment Inequality Models):** Suppose the identified set is the set of all solutions to a finite number of moment inequalities

$$\mathbb{E}(m(Z,\theta)) \geq 0,$$

where ψ is an M-dimensional vector of known functions, and $\theta_0 \in \mathbb{R}^K$.

- The identified set is thus given by

$$\Theta_I = \{\theta \in \mathbb{R}^K : \mathbb{E}(m(Z,\theta)) \geq 0\}.$$

- Consider the population objective function

$$Q(\theta) = \mathbb{E}(m(Z,\theta))'_- W \mathbb{E}(m(Z,\theta))_-,$$

where $(x)_-$ is component-wise non-positive part of x, and and W is a non-negative definite weight matrix.

Estimation: Criterion Function Approach IV

- Then the identified set is the set of all values of θ such that $Q(\theta) = 0$.
- A sample counterpart of this objective function would be

$$Q_n(\theta) = \left(\frac{1}{n} \sum_{i=1}^{n} m(Z_i, \theta) \right)_-^{\prime} W \left(\frac{1}{n} \sum_{i=1}^{n} m(Z_i, \theta) \right)_-$$

Estimation: Criterion Function Approach V

- First idea would be to estimate the identified set by

$$\tilde{\Theta}_I = \{\theta : Q_n(\theta) = 0\}.$$

- 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.
- **Intuition:** Consider the standard GMM case with equalities and overidentification.
  - Even if $\mathbb{E}(\psi(Z,\theta_0)) = 0$ the sample objective function will not be zero in finite samples in the case with over-identification.
- **Another Intuition:** Suppose that by construction $Q_n(\theta) \geq 0$ for all θ. If Q_n is not degenerate over the identified set we have that $\Pr(Q_n(\theta) > 0) > 0$ for $\theta \in \Theta_I$.
- As a result, $\tilde{\Theta}_I$ can e.g. be empty when Θ_I is not, even in large samples.

Estimation: Criterion Function Approach VI

- **A feasible approach:** Estimate Θ_I by the level set

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

  where $c_n \to 0$ at an appropriate rate.
- In most regular problems choosing $c_n = c\log(n)/n$ for some constant c is appropriate, and leads to an estimator of $\hat{\Theta}_I$ that is consistent in the Hausdorff norm.
- In particular, one can show that

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

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

Inference: General Principles I

- The role of inferential procedures is to quantify our uncertainty about our estimates due to using a finite data set.
- Inference in partially identified models is still an active area of research.
- There are many subtle issues that do not appear under point identification.
- We will start with a simple example to illustrate the problems.
- After that, we turn to are more general framework for inference.

Inference: General Principles II

- Suppose we want to compute a confidence set CS_n with level $1-\alpha$.
- **Problem:** What should the confidence set cover (asymptotically)?
- The entire identified set Θ_I?

$$\liminf_{n\to\infty} \Pr(\Theta_I \in CS_n) \geq 1-\alpha.$$

- Or the true parameter value θ_0?

$$\liminf_{n\to\infty} \Pr(\theta_0 \in CS_n) \geq 1-\alpha.$$

- Both approaches have been discussed in the literature, and both have their place in certain applications.
- The second notion is more in line with the traditional view of a confidence interval under point identification
- It is not clear why this intuition should be changed in partially identified models.

Inference: General Principles III

- Another problem under partial identification is the **uniform validity** of confidence sets.
- We might have that

$$\liminf_{n\to\infty} \Pr(\theta_0 \in CS_n) \geq 1-\alpha.$$

for one particular DGP.
- Still, for fixed n the probability $\Pr(\theta_0 \in CS_n)$ might depend a lot on the true DGP.
- It is thus useful to have confidence sets that satisfy

$$\liminf_{n\to\infty} \inf_{\text{valid DPGs}} \Pr(\theta_0 \in CS_n) \geq 1-\alpha.$$

- We will illustrate the last point in an example.

Inference: Moment Inequality Models I

- We now turn to inference on θ_0 in a general setting.
- We consider models that lead to a system of (unconditional) moment inequalities.
- Most of our examples can be cast in this framework.
- Parameter of interest is typically vector valued, and the identified set can have an arbitrary complicated form.
- Most of our discussion is based on work of Don Andrews with various co-authors.

Inference: Moment Inequality Models II

- **Model:** The true value θ_0 satisfies

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

- Here $m(\cdot,\theta) = (m_j(\cdot,\theta), j = 1,\dots,k)$ are known real-valued moment functions.
- θ_0 may or may not be identified by the moment conditions.
- **Aim:** Construct confidence sets for θ_0.
- Can be obtained by inverting a test $T_n(\theta)$ for testing $H_0 : \theta = \theta_0$:

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

- **Questions:** Which test statistic? Which critical value?
- We will discuss two types of test statistic and three types of critical values.

Inference: Moment Inequality Models III

- **General setup:** consider the sample moment functions

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

- Let $\hat{\Sigma}(\theta)$ be an estimator of the asymptotic variance, $\Sigma(\theta)$, of $n^{1/2}\bar{m}_n(\theta)$.
- For i.i.d. data we can take

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

Inference: Moment Inequality Models IV

- 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.

Inference: Moment Inequality Models V

Testing Functions

- **Example 1:** Modified Method of Moments (MMM). $S = S_1$ with

$$S_1(m, \Sigma) = \sum_{j=1}^{p} (m_j/\sigma_j)_-^2 + \sum_{j=p+1}^{p+v} (m_j/\sigma_j)^2$$

  where σ_j is jth diagonal element of Σ.
- The function S_1 yields a test statistic that gives positive weight to moment inequalities only when they are violated.
- This test is e.g. considered in Chernozhukov et al. (2007).

Inference: Moment Inequality Models VI

Testing Functions

- **Example 2:** Gaussian quasi-likelihood ratio (or minimum distance). $S = S_2$ with

$$S_2(m, \Sigma) = \inf_{t=(t_1, 0_v), t_1 \in \mathbb{R}^p_{+,[+\infty]}} (m-t)'\Sigma^{-1}(m-t)$$

- The reason we minimize over $t_1 \in \mathbb{R}^p_{+,[+\infty]}$ (and not just $\mathbb{R}^p_+$) is because for the asymptotic analysis we have to allow for $m_j = \infty$.
- This test is e.g. considered in Rosen (2008).
- Of course, other testing functions can also be considered.

Inference: Moment Inequality Models VII

Critical Values

- 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 $S(\Omega^{1/2}Z + h_1, \Omega)$, denoted by $c_{h_1}(1 - \alpha, \theta)$ or, at least, a consistent estimator of it.
- This requires knowledge of h_1, which cannot be estimated consistently.
- Different critical values are thus based on different approximations of $c_{h_1}(1 - \alpha, \theta)$.

Inference: Moment Inequality Models VIII

- **Approach 1:** Plug-in Asymptotics (PA).
- Can show 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 $S(\Omega^{1/2}Z, \Omega)$, denoted by $c_0(1-\alpha, \theta)$.
- PA critical values are defined as consistent estimators of $c_0(1-\alpha, \theta)$.
- 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 PA critical value is

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

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

Inference: Moment Inequality Models IX

Discussion of PA approach

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

$$\liminf_{n\to\infty} \inf_{\text{valid DGPs}} \Pr(\theta_0 \in CS_n^{PA}) \geq 1-\alpha$$

- PA critical values are conservative, since they are based on the least favorable case.
  - Coverage probability of resulting confidence sets is typically larger than $1-\alpha$.

Inference: Moment Inequality Models X

- **Approach 2:** Generalized Moment Selection (GMS); from Andrews and Soares (2010).
- Idea is to figure out which moment inequalities are binding from the data.
- For some $\kappa_n \to \infty$ at a suitable rate (e.g. $\kappa_n = (2\log(\log(n)))^{1/2}$) 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.

Inference: Moment Inequality Models XI

- GMS replaces h_1 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.
- Some common examples include:

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

- 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.

Inference: Moment Inequality Models XII

Discussion of GMS approach

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

$$\liminf_{n\to\infty} \inf_{\text{valid DGPs}} \Pr(\theta_0 \in CS_n^{GMS}) \geq 1-\alpha.$$

- GMS confidence sets are not asymptotically conservative under certain technical conditions:

$$\liminf_{n\to\infty} \inf_{\text{valid DGPs}} \Pr(\theta_0 \in CS_n^{GMS}) = 1-\alpha.$$

- Confidence sets have smaller volume than those based on PA.
- Confidence set depends on (arbitrary) choice of function φ.

Inference: Moment Inequality Models XIII

- **Approach 3:** Subsampling, (Politis and Romano, 1994).
- Subsampling tries to approximate the distribution of $T_n(\theta)$ directly.
- **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.
- **Subsampling** tries to mimic this infeasible approach:
  - 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 $T_n(\theta)$.
- Computationally intensive, but works in theory under very weak conditions.

Inference: Moment Inequality Models XIV

- Let b_n denote subsample size, which satisfies $b_n \to \infty$ and $b_n/n \to$ 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

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

Inference: Moment Inequality Models XV

Discussion of SS approach

- SS critical values require extensive computations.
- SS confidence sets are asymptotically valid in a uniform sense.

$$\liminf_{n\to\infty} \inf_{\text{valid DGPs}} \Pr(\theta_0 \in CS_n^{SS}) \geq 1-\alpha.$$

- SS confidence sets are not asymptotically conservative under certain technical conditions:

$$\liminf_{n\to\infty} \inf_{\text{valid DGPs}} \Pr(\theta_0 \in CS_n^{SS}) = 1-\alpha.$$

- 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.

Inference: Moment Inequality Models XVI

- 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
  - MMM test is more robust than QLR test,
  - PA 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.

Using the geometric structure to simplify the inference

- The main references are Beresteanu and Molinari (2008), Beresteanu, Molochanov and Molinari (2011), Bontemps, Magnac and Maurin (2012), Kaido and Santos (2013).
- When the set is convex, one can use the tools of the convex set theory (see Rockafellar, 1970) to propose simple estimators and testing strategies.
- Kaido and Santos (2013) estimate an efficiency bound and prove that the natural estimator of the support function is efficient
- It is also very simple to propose inference for subset of the vector of parameters.

No Moment Condition in Surplus

$$E(z^T(x\theta - y_c)) = E(z^T u(z)), u(z) \in \pm E(y_u - y_l|z)/2\}.$$

The identified set

$$\Theta_I = \{\theta : \theta = (E(z^T x))^{-1} E(z^T(y + u(z))), u(z) \in [-E(y_u - y_l|z)/2, E(y_u - y_l|z)/$$

is

- *Non empty*: θ^* corresponding to $u(z) = 0$ belongs to B.
- *Bounded*: $u(z)$ is uniformly bounded by bounds which are integrable (L_2).
- *Convex*: Moment conditions are linear and the interval containing $u(z)$ is convex.

Support Function

The dual to the indicator function of a convex set is called its support function, i.e.

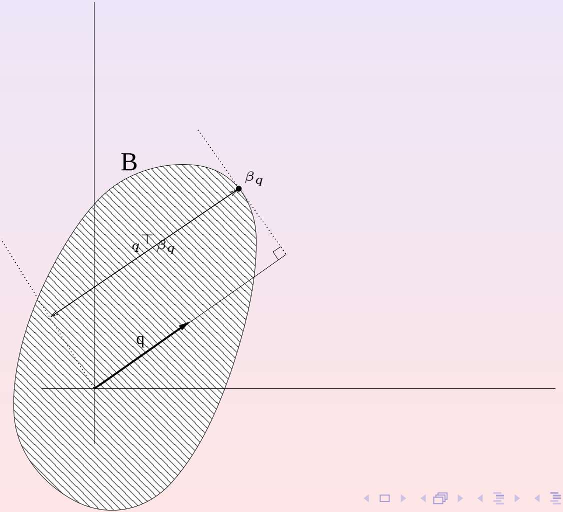
$$\delta^*(q \mid B) = \sup_{\theta \in B}(q^T\theta) \text{ for all directions, } q \text{ such that } \| q \| = 1.$$

A convex set can be fully described by its support function, (Rockafellar, 1970)

$$\theta \in B \Leftrightarrow \forall q, \| q \| = 1, q^T\theta \leq \delta^*(q \mid B).$$

The support function of a convex and bounded set is bounded and differentiable. Its derivative is continuous except at a countable number of points.

The support function



Hormander's embedding theorem

The Hausdorff distance between two sets A and B:

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

where $d(w,U) = \inf_{u\in U} d(w,u)$.
Isometry between the Hausdorff distance and the support function:

$$d_H(A,B) = \sup_{q\in\mathbb{S}} |\delta^*(q|A) - \delta^*(q|B)|,$$

where $\mathbb{S}$ is the unit sphere

$$\mathbb{S} = \{q : \parallel q \parallel = 1\}$$

Inference: The criterion view

Chernozhukov, Hong and Tamer (2007)

The identified set B is defined by a criterion:

$$Q(\theta) = 0 \Longleftrightarrow \theta \in B$$

A natural choice here is:

$$Q(\theta) = \int_{\mathbb{S}} (\delta^*(q \mid B) - q^T\theta)^2 \mathbf{1}\{\delta^*(q|B) < q^T\theta\} d\mu(q)$$

where $\mu(q)$ is a strictly positive measure on the unit sphere $\mathbb{S} \subset \mathbb{R}^p$.

The Identified Set: Projection in a Single Dimension

$$\Theta_I = \{\theta : \theta = (E(z^T x))^{-1} E(z^T (y_c + u(z))), u(z) \in -E(y_u - y_l|z)/2, E(y_u - y_l|z)/$$

Consider a direction q. We project the incomplete linear moment conditions

$$E(z^T (x\theta - y_c)) = E(z^T u(z))$$

onto direction q:

$$q^T \theta = q^T E(z^T x)^{-1} E\left(z^T (y_c + u(z))\right) = E\left(z_q (y_c + u(z))\right)$$

where

$$z_q = q^T E(z^T x)^{-1} z^T.$$

The Identified Set: Characterization

The support function $\delta^*(q \mid \theta_I) = \sup_{\theta \in \Theta_I}(q^T \theta)$ is the solution to a single-dimensional problem:

$$\sup_{u(z) \in [\underline{\Delta}(z), \overline{\Delta}(z)]} E(z_q(y + u(z)))$$

obtained using: $u_q(z) = \underline{\Delta}(z) + \left(\overline{\Delta}(z) - \underline{\Delta}(z)\right) \mathbf{1}\{z_q > 0\}$.
Result: The identified set Θ_I is defined by its support function

$$\delta^*(q \mid \Theta_I) = E(z_q(y + u_q(z))) = E(z_q w_q)$$

where w_q is an easy-to-construct variable:

$$w_q = \underline{y} + (\overline{y} - \underline{y})\mathbf{1}\{z_q > 0\}.$$

Remark:

$$\theta_q = E(z^T x)^{-1} E(z^T w_q)$$

Smoothness of the set

- If z has full support and his p.d.f. is strictly positive and continuous, the set Θ_I is smooth.
- If z the support is a subset of $\mathbb{R}$ and the p.d.f is strictly continuous, Θ_I has kinks.
- If z has mass points, Θ_I has exposed faces.
- If z is discrete Θ_I has kinks and exposed faces.

Asymptotic Properties I

$$\delta^*(q \mid \Theta_I) = E(z_q w_q)$$

where:

$$z_q = q^T E(z^T x)^{-1} z^T = q^T \Sigma^T z^T,$$
$$w_q = \underline{y} + \mathbf{1}\{z_q > 0\}(\overline{y} - \underline{y}).$$

Define an estimate $\hat{\Sigma}_n$ of Σ and define the empirical analogues:

$$z_{n,qi} = q^T \hat{\Sigma}_n^T z_i^T,$$
$$w_{n,qi} = \underline{y}_i + \mathbf{1}\{z_{n,qi} > 0\}(\overline{y}_i - \underline{y}_i).$$

The estimate of the support function is defined as:

$$\hat{\delta}_n^*(q \mid \Theta_I) = \frac{1}{n} \sum_{i=1}^{n} z_{n,qi} w_{n,qi}.$$

Asymptotic Properties II

If z has no mass point, the stochastic process

$$\tau_n(q) = \sqrt{n}(\hat{\delta}_n^*(q \mid \Theta_I) - \delta^*(q \mid \Theta_I)),$$

defined on the unit sphere, tends uniformly in distribution when n tends to ∞ to a Gaussian stochastic process, $d(q)$, such that:

$$E(d(q)) = 0$$

and the covariance operator is:

$$Cov(d(q)d(r)) = E(z_q z_r \varepsilon_q \varepsilon_r) - E(z_q \varepsilon_q)E(z_r \varepsilon_r).$$

$\varepsilon_q = w_q - x\theta_q$.

Tests

Here, we test $\{\theta_0\} \in \Theta_I$ using the support function:

$$\{\theta_0\} \in \Theta_I \Longleftrightarrow \forall q \in \mathbb{S},\ \delta^*(q \mid \Theta_I) - q^T \theta_0 \geq 0$$

For a frontier point, $\{\theta_0\} \in \partial\Theta_I$, there exists at least one direction q_0 for which the previous expression binds with equality:

$$\exists q_0 \in \mathbb{S},\ \delta^*(q_0 \mid \Theta_I) = q_0^T \theta_0$$

If Θ_I is strictly convex, q_0 is unique.

Test Procedure for $\theta_0 \in \Theta_I$ I

Based on the infimum of the following stochastic process on the unit sphere $\mathbb{S}$:

$$\sqrt{n} T_\infty(q; \theta_0) = \sqrt{n}(\delta^*(q \mid \Theta_I) - q^T \theta_0)$$

If $\theta_0 \in \Theta_I$ and q_0 is unique:

- $\sqrt{n} T_\infty(q; \theta_0) > 0$ for $q \neq q_0$,
- $\sqrt{n} T_\infty(q_0; \theta_0) = 0$.

We replace now $T_\infty(q; \theta_0)$ by its estimator $T_n(q; \theta_0)$ and base our test procedure on:

$$\sqrt{n} T_n(q; \theta_0) = \sqrt{n}(\hat{\delta}_n^*(q \mid \Theta_I) - q^T \theta_0)$$

Test Procedure for $\theta_0 \in \Theta_I$ II

- Search for **a** minimum q_n of

$$T_n(q;\theta_0) = \hat{\delta}_n^*(q \mid \Theta_I) - q^T\theta_0 \text{ on the unit sphere } \mathbb{S}.$$

- if q_0 is unique, q_n tends to q_0,

$$\sqrt{n}T_n(q_n;\theta_0) - \sqrt{n}T_n(q_0;\theta_0) \to 0.$$

-

$$\begin{aligned}\sqrt{n}T_n(q_0;\theta_0) &= \sqrt{n}(\hat{\delta}_n^*(q_0 \mid \Theta_I) - q_0^T\theta_0) \\ &= \sqrt{n}(\hat{\delta}_n^*(q_0 \mid \Theta_I) - \delta^*(q_0 \mid \Theta_I)) + \sqrt{n}(\delta^*(q_0 \mid \Theta_I)) - q_0^T\theta_0)\end{aligned}$$

- The first term converges to a Gaussian process with known variance V_{q_0},
- The second term is zero, positive or negative depending on the fact that $\theta_0 \in \partial\Theta_I$, $\theta_0 \in \Theta_I$, $\theta_0 \notin \Theta_I$.

Summary

- Search for **a** minimum q_n of

$$T_n(q;\theta_0) = \hat{\delta}_n^*(q \mid \Theta_I) - q^T\theta_0 \text{ on the unit sphere } \mathbb{S}.$$

- Compute the Studentized statistic of the minimum:

$$\xi_n(\theta_0) = \sqrt{n}\frac{T_n(q_n;\theta_0)}{\sqrt{\hat{V}_n}} = \sqrt{n}\frac{\min_q T_n(q;\theta_0)}{\sqrt{\hat{V}_n}}.$$

with $\hat{V}_n = V_{q_n} = Cov(d(q_n), d(q_n))$.
Then, if $\theta_0 \in \partial\Theta_I$,

$$\xi_n(\theta_0) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1),$$

if $\theta_0 \in int(\Theta_I)$,

$$\xi_n(\theta_0) \underset{n\to\infty}{\longrightarrow} +\infty$$

and if θ_0 does not belong to Θ_I,

$$\xi_n(\theta_0) \underset{n\to\infty}{\longrightarrow} -\infty.$$

- In many examples, point identification is ruled out because some information is missing.
- If one can bound this information, a set can be estimated.
- Despite the huge number of theoretical contributions, a few empirical applications only (see next course).
- There is still theoretical and empirical work to do.