Estimation and Confidence Regions for Parameter Sets in Econometric Models

Victor Chernozhukov, Han Hong and Elie Tamer

presented by

Yang, Yong Hyeon

KEYWORDS

- Criterion Based Estimation

  Want to minimize criterion function $Q(\theta)$

- Identified Set

$$\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\}$$

- Moment Condition (Equality and Inequality)

- Rate of Convergence and Confidence Region

MOMENT EQUALITY MODEL

- Consider

$$E[m_i(\theta)] = 0$$

- Criterion Function

$$Q(\theta) \equiv \left|\left| E[m_i(\theta)]' W^{1/2}(\theta) \right|\right|^2$$

- Identified Set

$$\begin{aligned} \Theta_I &= \{\theta \in \Theta : E[m_i(\theta)] = 0\} \\ &= \left\{ \theta \in \Theta : \left|\left| E[m_i(\theta)]' W^{1/2}(\theta) \right|\right|^2 = 0 \right\} \end{aligned}$$

- Sample Criterion Function

$$\begin{aligned} Q_n(\theta) &= \left|\left| E_n[m_i(\theta)]' W_n^{1/2}(\theta) \right|\right|^2 \\ &= \left|\left| \left(\sum_{i=1}^{n} m_i(\theta) \right)' W_n^{1/2}(\theta) \right|\right|^2 \end{aligned}$$

MOMENT INEQUALITY MODEL

- Consider

$$E[m_i(\theta)] \leq 0$$

- Criterion Function

$$\begin{aligned} Q(\theta) &\equiv \left|\left| E[m_i(\theta)]'_+ W^{1/2}(\theta) \right|\right|^2 \\ &= \left|\left| \max\left\{0, E[m_i(\theta)]'\right\} W^{1/2}(\theta) \right|\right|^2 \end{aligned}$$

- Identified Set

$$\begin{aligned} \Theta_I &= \{\theta \in \Theta : E[m_i(\theta)] \leq 0\} \\ &= \left\{\theta \in \Theta : \left|\left| E[m_i(\theta)]'_+ W^{1/2}(\theta) \right|\right|^2 = 0 \right\} \end{aligned}$$

- Sample Criterion Function

$$\begin{aligned} Q_n(\theta) &= \left|\left| E_n[m_i(\theta)]'_+ W_n^{1/2}(\theta) \right|\right|^2 \\ &= \left|\left| \left(\frac{1}{n} \sum_{i=1}^{n} m_i(\theta) \right)'_+ W_n^{1/2}(\theta) \right|\right|^2 \end{aligned}$$

RECALL POINT IDENTIFICATION (203B)

- IV model

$$E[z_i u_i] = 0$$

or equivalently

$$E[z_i(y_i - x_i'\theta)] = 0$$

- Identification condition is

$$\exists 1 \theta_0 \quad \text{s.t.} \quad E[z_i(y_i - x_i'\theta)] = 0$$

or equivalently $E[z_i x_i']$ is a full column rank matrix.

- A criterion function is

$$\begin{aligned} Q(\theta) &= \left|\left| E[z_i(y_i - x_i'\theta)]' W^{1/2}(\theta) \right|\right|^2 \\ &= E[z_i(y_i - x_i'\theta)]' W(\theta) E[z_i(y_i - x_i'\theta)] \end{aligned}$$

- A sample CF is

$$Q_n(\theta) = \left(\frac{1}{n} \sum_{i=1}^{n} z_i(y_i - x_i'\theta) \right)' W_n(\theta) \left(\frac{1}{n} \sum_{i=1}^{n} z_i(y_i - x_i'\theta) \right)$$

EXAMPLE 1

- Interval Data (Manski)

  The parameter of interest is $\theta = E[Y]$. Due to missing observation, we know

$$E[Y_{1i}] \leq \theta \leq E[Y_{2i}]$$

- Construct a criterion function

  Use

$$E\begin{bmatrix} Y_{1i} - \theta \\ \theta - Y_{2i} \end{bmatrix} \leq 0 \quad \text{and} \quad W(\theta) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

  so

$$\begin{aligned} Q(\theta) &= \left\| E\begin{bmatrix} Y_{1i} - \theta \\ \theta - Y_{2i} \end{bmatrix}_+^{\prime} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\|^2 \\ &= E\begin{bmatrix} Y_{1i} - \theta \\ \theta - Y_{2i} \end{bmatrix}_+^{\prime} E\begin{bmatrix} Y_{1i} - \theta \\ \theta - Y_{2i} \end{bmatrix}_+ = (E[Y_{1i}] - \theta)_+^2 + (E[Y_{2i}] - \theta)_-^2 \end{aligned}$$

  Corresponding sample CF is

$$Q_n(\theta) = (E_n[Y_{1i}] - \theta)_+^2 + (E_n[Y_{2i}] - \theta)_-^2$$

EXAMPLE 2

- Interval Regression

  Again the model looks like

$$E[Y_{1i}|X_i] \leq X_i'\theta \leq E[Y_{2i}|X_i]$$

- Use the moment inequality

$$E\left[\begin{array}{c} Y_{1i} - X_i'\theta \\ X_i'\theta - Y_{2i} \end{array}\middle| X_i\right] \leq 0$$

  or

$$E\left[\begin{array}{c} (Y_{1i} - X_i'\theta)Z_i \\ (X_i'\theta - Y_{2i})Z_i \end{array}\right] \leq 0$$

  to construct a criterion function

$$Q(\theta) = (E[(Y_{1i} - X_i'\theta)Z_i])_+^2 + (E[(Y_{2i} - X_i'\theta)Z_i])_-^2$$

  Corresponding sample CF is

$$Q_n(\theta) = (E_n[(Y_{1i} - X_i'\theta)Z_i])_+^2 + (E_n[(Y_{2i} - X_i'\theta)Z_i])_-^2$$

EXAMPLE 3

- Optimal Choice

  Agents choose between $D_i = 0, 1$. If an agent i chooses D_i, it gives more utility than $1 - D_i$.

$$\pi(W_i, D_i, \theta) + U_i \geq \pi(W_i, 1 - D_i, \theta) + V_i$$

  where U_i and V_i are independent with mean zero conditional on X_i.

- Use the moment inequality condition

$$E[\pi(W_i, 1 - D_i, \theta)|X_i] \leq E[\pi(W_i, D_i, \theta)|X_i]$$

  or

$$E\Big[\Big(\pi(W_i, 1 - D_i, \theta) - \pi(W_i, D_i, \theta)\Big) Z_i\Big] \leq 0$$

  to construct

$$Q(\theta) = \Big|\Big|\Big(\pi(W_i, 1 - D_i, \theta) - \pi(W_i, D_i, \theta)\Big) Z_i\Big|\Big|_+^2$$

EXAMPLE 4

- Structural Equations

  Consider the model

$$Y_i = \theta_0 + \theta_1 E + \theta_2 E^2 + \varepsilon$$

  We have the moment condition

$$E[(Y_i - X_i'\theta)Z_i] = 0$$

  where $Z = (1, I)$ and $X = (1, E, E^2)$, so $\dim(Z_i) \leq \dim(X_i)$.

- A criterion function is

$$Q(\theta) = \left|\left| E[(Y_i - X_i'\theta)Z_i] \right|\right|^2$$

  without a weight matrix.

- Sometimes we might be interested in

$$E\left[\left(\tau - 1(Y_i - X_i'\theta)\right)Z_i\right] = 0$$

  where τ is a parameter of interest.

NOTATION

- Hausdorff distance between sets

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

where a distance between a point and a set is usually defined as

$$d(a,B) = \inf_{b\in B} ||b-a||$$

- Contour set of level c

$$C_n(c) \equiv \{\theta \in \Theta : a_n Q_n(\theta) \leq c\}$$

where a_n is some normalizing sequence such as $a_n = n$.

- Inferential Statistic

$$\mathcal{C}_n \equiv \sup_{\theta\in\Theta_I} a_n Q_n(\theta)$$

ESTIMATION METHOD

- Estimation is Choosing c.

  Call such a number $\widehat{c}$. If we choose $\widehat{c}$, we have

  $$C_n(\widehat{c}) \equiv \{\theta \in \Theta : a_n Q_n(\theta) \leq \widehat{c}\}$$

  which is an estimator of

  $$\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\}$$

- $\widehat{c}$ might vary with n. More formally, it might depend on data.

- Consistency

  $$d_H\Big(C_n(\widehat{c}), \Theta_I\Big) \xrightarrow{p} 0$$

  How to choose $\widehat{c}$ so that $C_n(\widehat{c})$ is consistent?

CONFIDENCE REGION

- Test is Choosing c.

  If we choose a proper c, we have

$$C_n(c) \equiv \{\theta \in \Theta : a_n Q_n(\theta) \leq c\}$$

  such that

$$\Pr\Big(\Theta_I \subset C_n(c)\Big) = \alpha$$

- Note that ($\omega \in \Omega$ is an outcome)

$$\begin{aligned}\{\omega : \Theta_I \subset C_n(c)\} &= \{\omega : \forall \theta \in \Theta_I \ , \ a_n Q_n(\theta) \leq c\} \\ &= \left\{\omega : \sup_{\theta \in \Theta_I} a_n Q_n(\theta) \leq c\right\} \\ &= \{\omega : \mathcal{C}_n \leq c\}\end{aligned}$$

  So we only have to find c

$$\Pr(\mathcal{C}_n \leq c) = \alpha$$

  Need to know the distribution of $\mathcal{C}_n$.

EXAMPLE 1 (cont.)

- For consistent estimation, let $\widehat{c} = 0$. Recall

$$Q_n(\theta) = (\overline{Y}_1 - \theta)_+^2 + (\overline{Y}_2 - \theta)_-^2$$

So

$$a_n Q_n(\theta) \leq 0 \Leftrightarrow Q_n(\theta) = 0 \longrightarrow \widehat{\Theta}_I = \left[\overline{Y}_1, \overline{Y}_2\right]$$

This is consistent for $\left[E[Y_1], E[Y_2]\right]$

- For confidence region, let $a_n = n$.

$$\sqrt{n}\begin{pmatrix} \overline{Y}_1 - E[Y_1] \\ \overline{Y}_2 - E[Y_2] \end{pmatrix} \xrightarrow{d} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \stackrel{d}{=} N(0, \Omega)$$

So

$$\mathcal{C}_n = \sup_{\theta \in \left[E[Y_1], E[Y_2]\right]} n Q_n(\theta) \xrightarrow{d} \max\left[(W_1)_+^2, (W_2)_-^2\right] \equiv W$$

Find c such that (e.g. $\alpha = 0.95$)

$$\Pr(W \leq c) = \alpha$$

EXAMPLE 1 (cont.)

- Compare the result with Beresteanu and Molinari (2007).
  BM estimate Θ_I by

$$\widehat{\Theta}_I = \left[\overline{Y}_1, \overline{Y}_2\right]$$

  which is the same with CHT.

- For confidence region, BM used

$$\sqrt{n} \max \left\{ \left| \overline{Y}_1 - E[Y_1] \right| , \left| \overline{Y}_2 - E[Y_2] \right| \right\} \xrightarrow{d} \max\{|W_1|, |W_2|\} \equiv W_{BL}$$

  Looks similar. BM find δ such that

$$\Pr(W_{BL} \leq \delta) = \alpha$$

  and construct a confidence region

$$C_{BM} = \left\{ \theta : \overline{Y}_1 - \frac{\delta}{\sqrt{n}} \leq \theta \leq \overline{Y}_2 + \frac{\delta}{\sqrt{n}} \right\}$$

  CHT construct a confidence region

$$C_{CHT} = \{\theta : (\overline{Y}_1 - \theta)_+^2 + (\overline{Y}_2 - \theta)_-^2 \leq c\}$$

EXAMPLE 1 (cont.)

- What if we cannot derive the distribution of $\mathcal{C}_n$ analytically?
  CHT proposes subsampling (as well as simulation and bootstrap).

- A problem is that we need to know Θ_I to use subsampling.

$$\mathcal{C}_n = \sup_{\theta \in \left[E[Y_1], E[Y_2]\right]} n\left[(\overline{Y}_1 - \theta)_+^2 + (\overline{Y}_2 - \theta)_-^2\right]$$

Use a consistent estimator of Θ_I!

1. Using $\widehat{c} = 0$, estimate $\widehat{\Theta}_I = C_n(0) = [\overline{Y}_1, \overline{Y}_2]$.
2. Collect subsamples of size b.
3. For each subsample, calculate

$$\mathcal{C}_s^b = \sup_{\theta \in [\overline{Y}_1, \overline{Y}_2]} b\left[(\overline{Y}_1^b - \theta)_+^2 + (\overline{Y}_2^b - \theta)_-^2\right]$$

4. Find α-quantile of empirical distribution of $\mathcal{C}_s^b$ and define it as c. A confidence region is

$$C_n(c) = \{\theta \in \Theta : nQ_n(\theta) \leq c\}$$

CONSISTENCY (THEORY)

- Consistency

  Choose $\widehat{c}$ such that $\widehat{c} \geq \mathcal{C}_n$ w.p. 1 and $\frac{\widehat{c}}{a_n} \longrightarrow 0$. Then,

$$\lim_{n\to\infty} \Pr\left(\Theta_I \subseteq C_n(\widehat{c})\right) = 1$$

  and

$$d_H\left(C_n(\widehat{c}), \Theta_I\right) \xrightarrow{p} 0$$

  under the following condition.

  1. Θ is nonempty and compact.
  2. Q is l.s.continuous and $\inf_\Theta Q = 0$.
  3. Q_n is measurable.
  4. $\sup_\Theta (Q - Q_n)_+ = O_p(1/b_n)$ for some sequence $b_n \longrightarrow \infty$.
  5. $\sup_{\Theta_I} Q_n = O_p(1/a_n)$.

  In many cases, $b_n = \sqrt{n}$.

RATES OF CONVERGENCE (THEORY)

- Rates of convergence

$$d_H\Big(C_n(\widehat{c}), \Theta_I\Big) = O_p\left(\frac{1 \vee \widehat{c}}{a_n}\right)^{\frac{1}{\gamma}}$$

under the condition that there exist positive (δ, κ, γ) such that for any $\varepsilon \in (0,1)$, there are $(\kappa_\varepsilon, n_\varepsilon)$ satisfying for all $n > n_\varepsilon$,

$$Q_n(\theta) \geq \kappa\Big[d(\theta, \Theta_I) \wedge \delta\Big]^\gamma$$

uniformly on $\{\theta \in \Theta : d(\theta, \Theta_I) \geq (\kappa_\varepsilon / a_n)^{1/\gamma}\}$ with probability at least $1 - \varepsilon$.

- Notation

$$a \vee b \equiv \max(a, b)$$
$$a \wedge b \equiv \min(a, b)$$

- For example, in many cases $a_n = n$ and $\gamma = 2$. If we choose $\widehat{c}$ =constant, the rate of convergence is $\frac{1}{\sqrt{n}}$. If we choose $\widehat{c} \propto \log n$, the rate of convergence is $\sqrt{\log n / n}$.

CONSISTENCY WITH DEGENERACY

- Degeneracy condition is

  1. $\exists$ subsets $\Theta_n \subset \Theta$ such that $Q_n(\theta) = 0$ for all $\theta \in \Theta_n$, for each n and also $d_H(\Theta_n, \Theta_I) = \epsilon_n$ for some $\epsilon_n = o_p(1)$.
  2. ϵ_n is of stochastic order $a_n^{-1/\gamma}$.

- Let $\widehat{c} = O_p(1)$. Given consistency condition and degeneracy condition 1,

$$d_H\Big(C_n(\widehat{c}), \Theta_I\Big) \xrightarrow{p} 0$$

  and also rates of convergence condition and degeneracy condition imply

$$d_H\Big(C_n(\widehat{c}), \Theta_I\Big) = O_p\left(a_n^{-\frac{1}{\gamma}}\right)$$

- For example, with moment inequality conditions, we can find some ϵ-contraction as a candidate for such subsets, where

$$\epsilon\text{-expansion } \Theta_I^{\epsilon} \equiv \{\theta \in \Theta : d(\theta, \Theta_I) \leq \epsilon\}$$
$$\epsilon\text{-contraction } \Theta_I^{-\epsilon} \equiv \{\theta \in \Theta : d(\theta, \Theta \setminus \Theta_I) \geq \epsilon\}$$

EXAMPLE 1 AND 4 (cont.)

- Example 1

  $(\overline{Y}_1, \overline{Y}_2)$ is such a sequence of subsets satisfying degeneracy condition. So we can use $\widehat{c} = 0$ as we already saw. The rate of convergence is $\frac{1}{\sqrt{n}}$. Instead, we can set $\widehat{c} = \log n$. Consistency guaranteed and the rate of convergence is $\sqrt{\log n/n}$.

- Example 4

  When # of strong instruments $= k > d =$ # of endogenous variables,

$$J \equiv \inf_{\Theta} nQ_n \xrightarrow{d} \chi_{k-d}$$

Then

$$\Pr\Big(C_n(c) = \varnothing\Big) = \Pr(\chi_{k-d} > c) \equiv p > 0$$

So

$$\Pr\Big(d_H(C_n(c), \Theta_I) > 1\Big) = p$$

in the large samples. $C_n(c)$ is not consistent for Θ_I in this sense.

CONFIDENCE REGION (THEORY)

- Convergence Condition: $\mathcal{C}_n \xrightarrow{d} \mathcal{C}$
- Approximability Condition: Let Θ_n be such that $d_H(\Theta_n, \Theta_I) = o_p(a_n^{-1/\gamma})$. Define $\mathcal{C}'_n \equiv \sup_{\Theta_n} a_n Q_n(\theta)$. Then, for any $c \geq 0$,

$$\Pr(\mathcal{C}'_n \leq c) = \Pr(\mathcal{C} \leq c) + o(1)$$

- Theorem

  Given CONS, RATE, CONV and APPR conditions and also that

  1. data are iid,
  2. the size of subsample b is such that $b \to \infty$ and $\frac{b}{n} \to 0$, and
  3. $a_n \to \infty$

  subsampling method works well.

$$\lim_{n \to \infty} \Pr\left(\Theta_I \subseteq C_n(\widehat{c})\right) = \alpha$$

SUBSAMPLING

- Subsampling (simplified)

1. Collect subsamples of size b where .
2. Use appropriate $\widehat{c}_0$ to get consistent estimator $C_n(\widehat{c}_0)$.
3. Compute $\widehat{c}_l$ as a α quantile of the sample

$$\mathcal{C}_s^b = \sup_{\theta \in C_n(\widehat{c}_{l-1})} a_b Q_b(\theta)$$

4. Repeat step 3 for $l = 2, \cdots, L$.
5. Set $\widehat{c} = \widehat{c}_L$ and report $C_n(\widehat{c})$ as a confidence interval.

REST OF THE PAPER

- Section 3.5 : How to obtain asymptotics of $\mathcal{C}_n$.
- Section 4.1 : List primitive conditions for the case of moment equality condition case.
- Section 4.2 : List primitive conditions for the case of moment inequality condition case.
- Section 5 : List privimitive conditions for pointwise approach.
- In each of the above sections, simulation and bootstrap methods are explained.

OTHER PAPERS (recent econometric theory)

- Beresteanu and Molinari (2007) : random set theory, Wald based approach.
- Galichon and Henry (2006b) : criterion based, one-sided, simulation.
- Rosen (2006) : alternative criterion based, analytic.
- Bugni (2007) : the same method using bootstrap.
- some papers : projection methods
- other papers : global uniformity for subsampling

OTHER PAPERS (application)

- Application of Partial Identification?
  1. Varian (1982) : revealed preference
  2. Hansen, heaton, and Luttmer (1995) : asset pricing
  3. McFadden (2005) : revealed preference
  4. Blundell, Browning, and Crawford (2005) : revealed preference
  5. Haile and Tamer (2003) : auction
  6. Ciliberto and Tamer (2003) : market structure
  7. game theory, missing data, and stochastic dominance analysis.