EC824A
Extremum Estimators - Asymptotic Normality

Department of Economics
Michigan State University

Extremum Estimators – Distribution

- With consistency, would like to perform *inference*: *distributions* helpful for this
- Usually when econometricians mention *inference* they mean *distribution*
- We specify (and show) sufficient conditions for asymptotic normality
- Discuss assumptions in context of familiar estimators:
  1. ML
  2. Non-linear least squares (NLS)
  3. GMM
  4. MD
  5. TS
- Will maintain $Q_n(\theta)$ is smooth (twice differentiable) – non-smooth functions may be visited later in the course

Asymptotic Normality

- Now state sufficient conditions, later apply to example estimators

Assumption 1 (CF)

(i) *θ in interior of Θ*

(ii) *$G_n(\theta)$ twice continuously differentiable for neighbourhood $\Theta_0 \subset \Theta$ around θ_0*

(iii) $\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) \xrightarrow{d} N(0, \Omega_0)$

(iv) $\sup_{\theta\in\Theta_0}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta) - B(\theta)\right\| \xrightarrow{p} 0$ *for non-random square matrix function $B(\theta)$ that is continuous at θ_0 and $B_0 = B(\theta_0)$ nonsingular*

Assumption 2 (EE2)

(i) $\widehat{\theta}_n \xrightarrow{p} \theta_0$; (ii) $\frac{\partial}{\partial\theta}Q_n(\widehat{\theta}_n) = o_p(n^{-1/2})$.

Asymptotic Normality

Theorem 1

Assumptions CF and EE2 hold. Then $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, B_0^{-1}\Omega_0 B_0^{-1})$.

- The following is a useful result to prove the theorem

Lemma 2

Suppose (i) $\widehat{\theta}_n \xrightarrow{p} \theta_0$, *(ii)* $\sup_{\theta \in B(\theta_0,\epsilon)} \|L_n(\theta) - L(\theta)\| \xrightarrow{p} 0$ *for some* $\epsilon > 0$, *and (iii) non-stochastic function* $L(\theta)$ *continuous at* θ_0. *Then,* $\|L_n(\widehat{\theta}_n) - L(\theta_0)\| = o_p(1)$.

Proof of Lemma 2.

In class. ∎

Proof of Theorem 1.

In class. ∎

ML Estimator

- Recall $Q_n(\theta) := -\frac{1}{n}\sum_{i=1}^{n} \log f(W_i, \theta)$. Then,

$$\frac{\partial}{\partial\theta} Q_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(W_i, \theta)$$

$$\frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta'} \log f(W_i, \theta),$$

$$\Omega_0 = \mathbb{E}\left[\frac{\partial}{\partial\theta} \log f(W_i, \theta_0) \frac{\partial}{\partial\theta'} \log f(W_i, \theta_0)\right], \text{ and}$$

$$B(\theta_0) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta\partial\theta'} \log f(W_i, \theta_0)\right]$$

ML Estimator

- CF(ii) holds if $f(w, \theta)$ twive continuously differentiable in θ on some neighbourhood $\Theta_0 \subset \Theta$ of θ_0 for all w in support $\mathcal{W}$ of W_i.
- CF(iii) holds by CLT for i.i.d. random vectors with finite second moments if,

$$\mathbb{E}\left[\frac{\partial}{\partial\theta} Q_n(\theta_0)\right] = 0 \text{ and } \mathbb{E}\left[\left\|\frac{\partial}{\partial\theta} \log f(W_i, \theta)\right\|^2\right] < \infty. \tag{1}$$

- The first condition holds by FOC with θ_0 as minimiser, provided θ_0 is in the interior of Θ

$$0 = \frac{\partial}{\partial\theta} Q(\theta_0) = -\frac{\partial}{\partial\theta}\mathbb{E}[\log f(W_i, \theta_0)] = -\mathbb{E}\left[\frac{\partial}{\partial\theta} \log f(W_i, \theta_0)\right] = \mathbb{E}\left[\frac{\partial}{\partial\theta} Q_n(\theta_0)\right]$$

when $\mathbb{E}$ and $\frac{\partial}{\partial\theta}$ can inter-change (DCT). The equality holds by definition of θ_0 regardless of whether model correctly specified.

ML Estimator

- Second condition in (1) equivalent to Fisher information matrix at θ_0 well defined:

$$\mathcal{I}_0 = \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(W_i,\theta)\frac{\partial}{\partial\theta'}\log f(W_i,\theta)\right]$$

- "Information matrix equality" is $B_0 = \Omega_0$.
- This holds if **parametric model correctly specified** and order of differentiation and integration can switch for $B_0 = B(\theta_0)$
- In this case ("$\sim$": is distributed as),

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, B_0^{-1}\Omega_0 B_0^{-1}) \sim N(0, B_0^{-1})$$

ML Estimator

- CF(iv) holds if
  1. U-WCON over $\theta \in \Theta_0$ holds for

$$\left\{ -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta \partial \theta'} \log f(W_i, \theta) : n \geqslant 1 \right\}$$

  2. B_0 non-singular
  3. $B(\theta)$ continuous at θ_0.
- Continuity of $B(\theta)$ at θ_0 holds by Lemma 1 lecture 1 provided $\frac{\partial^2}{\partial \theta \partial \theta'} \log f(w, \theta)$ continuous in θ on Θ_0 for all $w \in \mathcal{W}$, and $\mathbb{E}\left[\sup_{\theta \in \Theta_0} \left\| \frac{\partial^2}{\partial \theta \partial \theta'} \log f(W_i, \theta) \right\|\right] < \infty$.

NLS Estimator

- $Q_n(\theta) := \frac{1}{n}\sum_{i=1}^n (Y_i - g(X_i,\theta))^2/2$. Then,

$$\frac{\partial}{\partial\theta} Q_n(\theta) = -\frac{1}{n}\sum_{i=1}^n (Y_i - g(X_i,\theta))\frac{\partial}{\partial\theta} g(X_i,\theta)$$

$$\frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta) = \frac{1}{n}\sum_{i=1}^n \left(\frac{\partial}{\partial\theta} g(X_i,\theta)\frac{\partial}{\partial\theta'} g(X_i,\theta) - (Y_i - g(X_i,\theta))\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta'} g(X_i,\theta)\right),$$

$$\Omega_0 = \mathbb{E}\left[(Y_i - g(X_i,\theta))^2 \frac{\partial}{\partial\theta} g(X_i,\theta)\frac{\partial}{\partial\theta'} g(X_i,\theta)\right], \text{ and}$$

$$B(\theta) = \mathbb{E}\left[\frac{\partial}{\partial\theta} g(X_i,\theta)\frac{\partial}{\partial\theta'} g(X_i,\theta)\right] - \mathbb{E}\left[(Y_i - g(X_i,\theta))\frac{\partial}{\partial\theta\partial\theta'} g(X_i,\theta)\right]$$

NLS Estimator

- CF(ii) holds if $g(x,\theta)$ twice differentiable on some neighbourhood $\Theta_0 \subset \Theta$ of θ_0 for all x in support $\mathcal{X}$ of X_i
- Recall $\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta) = -n^{-1/2}\sum_{i=1}^n (Y_i - g(X_i,\theta))\frac{\partial}{\partial\theta}g(X_i,\theta)$. CF(iii) holds by CLT for i.i.d. random vectors with finite second moment provided,

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}Q_n(\theta_0)\right] = 0, \text{ and } \mathbb{E}\left[(Y_i - g(X_i,\theta))^2 \left\|\frac{\partial}{\partial\theta}g(X_i,\theta)\right\|^2\right] < \infty.$$

- As before, for first condition holds by FOC since θ_0 is minimiser (if in interior of Θ)
- Then, can show,

$$0 = \frac{\partial}{\partial\theta}Q(\theta_0) = \cdots = \mathbb{E}\left[\frac{\partial}{\partial\theta}Q_n(\theta_0)\right]$$

provided (as before) $\mathbb{E}$ and $\frac{\partial}{\partial\theta}$ interchangeable (DCT). Again, holds by definition of θ_0, whether or not model correctly specified.

NLS Estimator

- CF(iv) holds by Theorem 3 (previous lecture) provided
  1. $\frac{\partial}{\partial\theta}g(x,\theta)$ and $\frac{\partial^2}{\partial\theta\partial\theta'}g(x,\theta)$ continuous in θ on Θ_0 $\forall x \in \mathcal{X}$,
  2. $$\mathbb{E}\left[\sup_{\theta\in\Theta_0}\left\|\frac{\partial}{\partial\theta}g(X_i,\theta)\right\|^2 + \left\|(Y_i - g(X_i,\theta))\frac{\partial^2}{\partial\theta\partial\theta'}g(X_i,\theta)\right\|\right] < \infty,$$
  3. Θ_0 compact,
  4. B_0 nonsingular.

NLS Estimator

- When model correctly specified ($\mathbb{E}[U_i|X_i] = 0 \;\; a.s.$)
- Asymptotic covariance of $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ is $B_0^{-1}\Omega_0 B_0^{-1}$, where,

$$\Omega_0 = \mathbb{E}\left[\sigma^2(X_i)\frac{\partial}{\partial\theta}g(X_i,\theta)\frac{\partial}{\partial\theta'}g(X_i,\theta)\right]$$
$$B_0 = \mathbb{E}\left[\frac{\partial}{\partial\theta}g(X_i,\theta)\frac{\partial}{\partial\theta'}g(X_i,\theta)\right]$$

- If $\{U_i : i \geqslant 1\}$ homoskedastic, $\sigma^2(X_i) = \sigma^2$, then $\Omega_0 = \sigma^2 B_0$, and AsyVar is $\sigma^2 B_0^{-1}$.

GMM Estimator

- $Q_n(\theta) := \|A_n \frac{1}{n}\sum_{i=1}^n m(W_i,\theta)\|^2/2$ for $\theta \in \Theta \subset \mathbb{R}^d$, $m \mapsto \mathbb{R}^k$. Then,

$$\frac{\partial}{\partial\theta} Q_n(\theta) = \left(\frac{1}{n}\sum_{i=1}^n \frac{\partial m(W_i,\theta)}{\partial\theta'}\right)' A_n' A_n \left(\frac{1}{n}\sum_{i=1}^n m(W_i,\theta)\right)$$

$$\left[\frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta)\right]_{\ell,j} = \left(\frac{1}{n}\sum_{i=1}^n \frac{\partial m(W_i,\theta)}{\partial\theta_\ell}\right)' A_n' A_n \left(\frac{1}{n}\sum_{i=1}^n \frac{\partial m(W_i,\theta)}{\partial\theta_j}\right) + \left(\frac{1}{n}\sum_{i=1}^n \frac{\partial^2 m(W_i,\theta)}{\partial\theta_\ell\partial\theta_j}\right)' A_n' A_n \left(\frac{1}{n}\sum_{i=1}^n m(W_i,\theta)\right)$$

GMM Estimator

- Also have for $\ell, j = 1, \ldots, d$. We also have

$$\begin{aligned}
\Gamma_0 &\equiv \mathbb{E}\left[\frac{\partial m(W_i, \theta_0)}{\partial \theta'}\right] \text{ and } A_n \xrightarrow{p} A \\
V_0 &\equiv \mathbb{E}\left[m(W_i, \theta_0) m(W_i, \theta_0)'\right] \\
[B(\theta)]_{\ell,j} &= \mathbb{E}\left[\frac{\partial m(W_i, \theta)}{\partial \theta_\ell}\right]' A'A \cdot \mathbb{E}\left[\frac{\partial m(W_i, \theta)}{\partial \theta_j}\right] \\
&\quad + \mathbb{E}\left[\frac{\partial^2 m(W_i, \theta)}{\partial \theta_\ell \partial \theta_j}\right]' A'A \cdot \mathbb{E}\left[m(W_i, \theta)\right] \\
\Omega_0 &= \Gamma_0' A' A V_0 A' A \Gamma_0
\end{aligned}$$

- Note that we have $B_0 = \Gamma_0' A' A \Gamma_0$ since $\mathbb{E}\left[m(W_i, \theta_0)\right] = 0$. It follows that

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, B_0^{-1} \Omega_0 B_0^{-1})$$

where $B_0^{-1} \Omega_0 B_0^{-1} = \left(\Gamma_0' A' A \Gamma_0\right)^{-1} \Gamma_0' A' A V_0 A' A \Gamma_0 \left(\Gamma_0' A' A \Gamma_0\right)^{-1}$.

GMM Estimator

- CF (ii) holds if $m(w, \theta)$ is twice continuously differentiable for all $w \in \mathcal{W}$
- CF (iii) holds by (i) CLT applied to $\frac{1}{\sqrt{n}} \Sigma_{i=1}^{n} m(W_i, \theta_0)$ and (ii) WLLN applied to $\frac{1}{n} \Sigma_{i=1}^{n} \frac{\partial m(W_i, \theta_0)}{\partial \theta'}$, and (iii) $A_n \xrightarrow{p} A$. The CLT and U-WCON hold under:

$$\mathbb{E}[\| m(W_i, \theta_0) \|^2] < \infty \text{ and } \mathbb{E}\left[\left\| \frac{\partial m(W_i, \theta_0)}{\partial \theta'} \right\| \right] < \infty,$$

- CF (iv) holds if (i) $\left\{ \frac{\partial^2 m(W_i, \theta)}{\partial \theta_\ell \partial \theta_j} : i \geqslant 1 \right\}$ for $l, j = 1, \ldots, d$, $\left\{ \frac{\partial m(W_i, \theta)}{\partial \theta'} : i \geqslant 1 \right\}$, and $\{ m(W_i, \theta) : i \geqslant 1 \}$ satisfy Assumption SE1, (ii) Γ_0 full rank, (iii) A nonsingular, and (iv) $E\left[\frac{\partial^2 m(W_i, \theta)}{\partial \theta_\ell \partial \theta_j} \right]'$ for $l, j = 1, \ldots, d$, $E\left[\frac{\partial m(W_i, \theta)}{\partial \theta_\ell} \right]$, and $E\left[m(W_i, \theta) \right]$ are continuous at θ_0.
- (These are same conditions as Assumption SE1 from Lecture 1)

MD Estimator

- Recall that $Q_n(\theta) = \|A_n(\hat{\pi}_n - g(\theta))\|^2/2$. We have

$$
\begin{aligned}
\frac{\partial}{\partial\theta}Q_n(\theta) &= \left(\frac{\partial}{\partial\theta'}g(\theta)\right)' A_n'A_n\left(\widehat{\pi}_n - g(\theta)\right), \\
\left[\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta)\right]_{\ell,j} &= \frac{\partial}{\partial\theta_\ell}g(\theta)'A_n'A_n\frac{\partial}{\partial\theta_j}g(\theta) - \frac{\partial^2}{\partial\theta_\ell\partial\theta_j}g(\theta)'A_n'A_n\left(\widehat{\pi}_n - g(\theta)\right)
\end{aligned}
$$

  for $\ell, j = 1, \ldots, d$.

- Assume

$$
\sqrt{n}(\widehat{\pi}_n - \pi_0) \xrightarrow{d} N(0, V_0) \text{ and } A_n \xrightarrow{p} A. \tag{2}
$$

  This assumption can be established using Theorem 1 if $\widehat{\pi}_n$ is an extremum estimator.

MD Estimator

- Also have,

$$\Omega_0 = \Gamma_0' A' A V_0 A' A \Gamma_0, \text{ where } \Gamma_0 = \frac{\partial}{\partial \theta'} g(\theta), \text{ and}$$

$$[B(\theta)]_{\ell,j} = \frac{\partial}{\partial \theta_\ell} g(\theta)' A' A \frac{\partial}{\partial \theta_j} g(\theta) - \frac{\partial^2}{\partial \theta_\ell \partial \theta_j} g(\theta)' A' A \left(\pi_0 - g(\theta)\right) \text{ for } \ell, j = 1, \ldots, d.$$

- CF(ii) holds if $g(\theta)$ is twice continuously differentiable in θ on some neighborhood $\Theta_0 \subset \Theta$ of θ_0
- CF(iii) holds by (2) provided the restriction $\pi_0 = g(\theta_0)$ holds (where θ_0 is the probability limit of $\widehat{\theta}_n$. Note that $B(\theta_0)$ simplifies to $\Gamma_0' A' A \Gamma_0$ under $\pi_0 = g(\theta_0)$.
- CF (iv) trivially holds under the assumptions given above, provided that Γ_0 and A have full rank.

TS Estimator

- Recall $Q_n(\theta) = \|A_n G_n(\theta, \hat{\tau}_n)\| / 2$, such that,

$$
\begin{aligned}
\frac{\partial}{\partial \theta} Q_n(\theta) &= \left(\frac{\partial G_n(\theta, \widehat{\tau}_n)}{\partial \theta'} \right)' A_n' A_n G_n(\theta, \widehat{\tau}_n) \\
\left[\frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta) \right]_{\ell, j} &= \left(\frac{\partial}{\partial \theta_\ell} G_n(\theta, \widehat{\tau}_n) \right)' A_n' A_n \left(\frac{\partial}{\partial \theta_j} G_n(\theta, \widehat{\tau}_n) \right) \\
&\quad + \left(\frac{\partial^2}{\partial \theta_\ell \partial \theta_j} G_n(\theta, \widehat{\tau}_n) \right)' A_n' A_n G_n(\theta, \widehat{\tau}_n) \text{ for } \ell, j = 1, \ldots, d.
\end{aligned}
$$

- Left as exercise to show Assumption CF holds here.
- Focus on asymptotic covariance matrix

TS Estimator

- Assume

$$\sqrt{n}\begin{pmatrix} G_n(\theta_0,\tau_0) \\ \widehat{\tau}_n-\tau_0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{matrix} V_{10} & V_{20} \\ V_{20}' & V_{22} \end{matrix}\right), \tag{3}$$

$$\frac{\partial G_n(\theta_0,\widehat{\tau}_n)}{\partial\theta'} \xrightarrow{p} \frac{\partial G(\theta_0,\tau_0)}{\partial\theta'} = \Gamma_0,\ A_n \xrightarrow{p} A,$$

$$\frac{\partial G_n(\theta_0,\widehat{\tau}_n^*)}{\partial\tau'} \xrightarrow{p} \frac{\partial G(\theta_0,\tau_0)}{\partial\tau'} = \Lambda_0 \text{ for any } \widehat{\tau}_n^* \to \tau_0$$

$$\sup_{\theta\in\Theta_0}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta)-B(\theta)\right\| \xrightarrow{p} 0, \text{ where}$$

$$[B(\theta)]_{\ell,j} = \frac{\partial}{\partial\theta_\ell}G(\theta,\tau_0)'A'A\frac{\partial}{\partial\theta_j}G(\theta,\tau_0)+\frac{\partial^2}{\partial\theta_\ell\partial\theta_j}G(\theta,\tau_0)'A'AG(\theta,\tau_0)$$

for $\ell, j = 1,\ldots,d$

- Obtain $B(\theta_0)=\Gamma_0'AA'\Gamma_0$ if $G(\theta_0,\tau_0)=0$.

TS Estimator

- For asymptotic distribution of $\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0)$, carry out element-by-element mean value expansions of $\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0)$ around τ_0 and use the assumptions of (3)

$$
\begin{aligned}
\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) &= \left(\frac{\partial G_n(\theta_0,\widehat{\tau}_n)}{\partial\theta'}\right)' A_n' A_n \sqrt{n} G_n(\theta_0,\widehat{\tau}_n) \\
&= \left(\Gamma_0 + o_p(1)\right)' A_n' A_n \left(\sqrt{n} G_n(\theta_0,\tau_0) + \frac{\partial G_n(\theta_0,\widehat{\tau}_n^*)}{\partial\tau'}\sqrt{n}\left(\widehat{\tau}_n - \tau_0\right)\right) \\
&= \left(\Gamma_0 + o_p(1)\right)' A_n' A_n \left(I_k : \Lambda_0 + o_p(1)\right)\sqrt{n}\begin{pmatrix} G_n(\theta_0,\tau_0) \\ \widehat{\tau}_n - \tau_0 \end{pmatrix} \\
&\xrightarrow{d} \Gamma_0' A' A \left(Z_1 + \Lambda_0 Z_2\right) \sim N(0,\Omega_0), \text{ where} \\
\Omega_0 &= \Gamma_0' A' A \left(V_{10} + \Lambda_0 V_{20}' + V_{20}\Lambda_0' + \Lambda_0 V_{22}\Lambda_0'\right) A' A \Gamma_0
\end{aligned}
$$

  where $\widehat{\tau}_n^*$ lies between $\widehat{\tau}_n$ and τ_0. Denote $B_0 := \Gamma_0' A' A \Gamma_0$

- Conclude $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, B_0^{-1}\Omega_0 B_0^{-1})$