EC824A
Extremum Estimators

Department of Economics
Michigan State University

Extremum Estimators

- "Extremum": min or max of some criterion
- Consistency: given the estimator defined by the extremum of the criterion function, we show convergence to the true parameter of interest
- Convergence: limit of estimator when model misspecified

Examples – ML

- Maximum Likelihood (ML)
- Data $\{W_i : i \leqslant n\}$ i.i.d. density $f(w,\theta)$
- Likelihood function: $\prod_{i=1}^n f(W_i,\theta)$; Log-likelihood function: $n^{-1}\sum_{i=1}^n \log f(W_i,\theta)$.
- ML minimises over $\theta \in \Theta$:

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

- If $W_i = (Y_i', X_i')$. Conditional $f(y|x,\theta)$, marginal $g(x)$, then $f(w,\theta) = f(y|x,\theta)g(x)$,

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

- Second term independent of θ, so can ignore:

$$Q_n(\theta) \propto -\frac{1}{n}\sum_{i=1}^n \log f(Y_i|X_i,\theta)$$

Examples – LS

- $W_i = (Y_i', X_i')$ i.i.d. and nonlinear regression with $\mathbb{E}[U_i|X_i]$,

$$Y_i = g(X_i, \theta_0) + U_i$$

- LS estimator $\widehat{\theta}_n$ minimises over $\theta \in \Theta$,

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

Examples – GMM

- $W_i = (Y_i', X_i')$ i.i.d.
- Moment conditions hold, where $m(x, \theta) \in \mathbb{R}^k$ for $k \geqslant dim(\Theta)$,

$$\mathbb{E}[m(W_i, \theta_0)] = 0$$

- For $k \times k$ random weight matrix A_n, GMM estimator $\widehat{\theta}_n$ minimises

$$Q_n(\theta) = \left\| A_n \frac{1}{n} \sum_{i=1}^{n} m(W_i, \theta) \right\|^2 / 2$$

over $\theta \in \Theta$, where $\| \cdot \|$ is Euclidean norm

Examples – MD

- $\widehat{\pi}_n$ unrestricted estimator of k-vector parameter π_0
- π_0 known function of d-vector parameter θ_0, $d \leqslant k$

$$\pi_0 = g(\theta_0)$$

- A_n $k \times k$ random weight matrix, over $\theta \in \Theta$ MD estimator $\widehat{\theta}_n$ minimises,

$$Q_n(\theta) = \|A_n(\widehat{\pi}_n - g(\theta))\|^2/2$$

- E.g. [Chamberlain (1982, 1984)] , for α_i correlated with x_{it}

$$y_{it} = \alpha_0 + x_{it}\beta + \alpha_i + u_{it}$$

- Linear projection of α_i on x_{it} written as,

$$\alpha_i = \lambda_0 + x_{i1}\lambda_1 + \cdots + x_{iT}\lambda_T + v_i$$

Examples – MD

- Plug-in to y

$$\begin{aligned} y_{it} &= \psi + x_{i1}\lambda_1 + \cdots + x_{it}(\beta + \lambda_t) + \cdots + x_{iT}\lambda_T + r_{it} \\ &= \pi_{t0} + x_{i1}\pi_{t1} + \cdots + x_{iT}\pi_{tT} + r_{it} \end{aligned}$$

  where $\psi = \alpha_0 + \lambda_0$, $r_{it} = v_i + u_{it}$.
- $\pi_t := \{\pi_{t0}, \ldots, \pi_{tT}\}'$ can be estimated by cross sectional regression for time t.
- If θ collects parameters $\alpha_0, \beta, \lambda_0, \ldots, \lambda_T$, then $\pi = H\theta$

Examples – TS

- $\{W_i : i \leqslant n\}$ i.i.d., $\widehat{\tau}_n$ preliminary estimator of τ_0
- Random k-vector $G_n(\theta, \tau) \in (-\epsilon, \epsilon)$ small ϵ when $\theta = \theta_0$, $\tau = \tau_0$, and $n \to \infty$
- E.g. GMM, $G_n(\theta, \tau) = \frac{1}{n} \sum_{i=1}^n g(W_i, \theta)$; in MD, $G_n(\theta, \tau) = \widehat{\pi}_n(\tau) - g(\theta, \tau)$
- A_n $k \times k$ random weight matrix. TS estimator $\widehat{\theta}_n$ minimises over $\theta \in \Theta$,

$$Q_n(\theta) = \|A_n G_n(\theta, \widehat{\tau})\|^2/2$$

- E.g. consider nonlinear model, with $\mathbb{E}[U_i^2 | X_i = x] = \sigma(x, \tau_0)$,

$$Y_i = g(X_i, \theta_0) + U_i$$

- Suppose $\widehat{\tau}_n$ consistent for τ_0, feasible WLS is,

$$\widehat{\theta}_{WLS} = \operatorname*{argmin}_{\theta \in \Theta} \frac{1}{2n} \sum_{i=1}^n \frac{(y_i - g(x_i, \theta))^2}{\sigma(x_i, \widehat{\tau}_n)}$$

Examples – TS

- FOCs

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}g(x_i,\theta)\frac{y_i-g(x_i,\theta)}{\sigma(x_i,\widehat{\tau}_n)}=0$$

- A two step estimator can be defined as

$$\widehat{\theta}_n=\underset{\theta\in\Theta}{\operatorname{argmin}}\,\|G_n(\theta,\widehat{\tau}_n)\|^2/2$$

with

$$G_n(\theta,\widehat{\tau}_n)=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}g(x_i,\theta)\frac{y_i-g(x_i,\theta)}{\sigma(x_i,\widehat{\tau}_n)}$$

Population Criterion Function

- To derive consistency consider population criterion function for each example
- ML: $Q(\theta) = -\mathbb{E}[l(W_i, \theta)] = -\mathbb{E}[\log f(W_i, \theta)]$, expections over true D.G.P.
- LS: $Q(\theta) = \mathbb{E}[(Y_i - g(X_i, \theta))^2/2]$
- GMM: $Q(\theta) = \|A \cdot \mathbb{E}[m(W_i, \theta)]\|^2/2$ for $A_n \xrightarrow{p} A$
- MD: $Q(\theta) = \|A \cdot (\pi_0 - g(\theta))\|^2/2$ for $A_n \xrightarrow{p} A$
- TS: $Q(\theta) = \|A \cdot G(\theta, \tau_0)\|^2/2$ for $A_n \xrightarrow{p} A$, $\widehat{\tau}_n \xrightarrow{p} \tau_0$, and $G_n(\theta, \tau) \xrightarrow{p} G(\theta, \tau)$

Consistency

- Need to confirm two conditions:
  1. Identifies true parameter
  2. Uniform convergence of sample criterion function to population criterion function
- $B(\theta_0, \epsilon)$ is ϵ-neighbourhood around θ_0: $B(\theta_0, \epsilon) := \{\theta \in \Theta : \|\theta - \theta_0\| < \epsilon\}$

Assumption 1 (EE)

$\widehat{\theta}_n \in \Theta$ *and* $Q_n(\widehat{\theta}_n) \leqslant \inf_{\theta\in\Theta} Q_n(\theta) + o_p(1)$

Assumption 2 (U-WCON)

$\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| = o_p(1)$ *for non-stochastic function* $Q : \Theta \mapsto \mathbb{R}$

Assumption 3 (ID)

$\forall\epsilon > 0$, $\inf_{\theta\in\Theta\setminus B(\theta_0,\epsilon)} Q(\theta) > Q(\theta_0)$

Consistency

Theorem 1

Let Assumptions EE, U-WCON, and ID hold. Then $\widehat{\theta}_n = \theta_0 + o_p(1)$.

Proof.

In class. ∎

Consistency

- Assumptions ID and U-WCON too strong to impose generally: verified per estimator
- High-level consistency very useful: only need to verify ID and U-WCON now
- Now to verify conditions for ID and U-WCON for some estimators

Sufficient conditions for ID

$$\forall \epsilon > 0, \inf_{\theta \in \Theta \setminus B(\theta_0, \epsilon)} Q(\theta) > Q(\theta_0)$$

- Condition for Assumption ID is that θ_0 uniquely minimises $Q(\theta)$ over Θ.

Assumption 4 (ID1)

(i). Θ *is compact*

(ii). $Q(\theta)$ *is continuous,*

(iii). θ_0 *uniquely minimises* $Q(\theta)$ *over* Θ

- (i) "simple" to verify: stipulated for regularity. E.g. $|\mu|, \sigma < \infty$, in $N(\mu, \sigma)$
- (iii) usually from stronger sufficient conditions: *essentially* assumed

Sufficient conditions for ID

- ID1.(ii)

Lemma 1

$m(w,\theta)$ is continuous in θ for all w in support $\mathcal{W}$ or r.v. W_i, $m(\cdot,\theta)$ integrable for each fixed $\theta \in \Theta$, and $\mathbb{E}[\sup_{\theta\in\Theta} \|m(W_i,\theta)\|] < \infty$. Then, $\mathbb{E}[\|m(W_i,\theta)\|]$ continuous in $\theta \in \Theta$.

Proof.

In class. ∎

Sufficient conditions for ID

- Use Lemma 1 to show ID1.(ii) holds for:
  1. ML if $f(w,\theta)$ continuous in $\theta \in \Theta$ for all w supported on distribution of W_i and $\mathbb{E}[\sup_{\theta\in\Theta} \| \log f(W_i,\theta)\|] < \infty$
  2. LS if $g(x,\theta)$ continuous in $\theta \in \Theta$ for all x supported on distribution of X_i and $\mathbb{E}[\sup_{\theta\in\Theta}(Y_i - g(X_i,\theta))^2] < \infty$
  3. GMM if $m(w,\theta)$ continuous in $\theta \in \Theta$ for all w supported on distribution of W_i and $\mathbb{E}[\sup_{\theta\in\Theta} \|mW_i,\theta)\|] < \infty$
  4. MD if $g(\theta)$ continuous in $\theta \in \Theta$
  5. TS if $G(\theta,\tau)$ continuous in $\theta \in \Theta$

Sufficient conditions for ID - ML

- ID1.(iii) [θ_0 uniquely minimises $Q(\theta)$ over Θ] in ML
- Let $f(w,\theta_0)$ be true distribution of W in parametric family: $\{f(w,\theta) : \theta \in \Theta\}$

$$Q_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n}[\log f(W_i,\theta)] \text{ and } Q(\theta) = -\mathbb{E}[\log f(W_i,\theta)]$$

- Consider,

$$\begin{aligned} Q(\theta_0) - Q(\theta) &= -\mathbb{E}[\log f(W_i,\theta_0)] + \mathbb{E}[\log f(W_i,\theta)] = \dots \text{ (in class)} \\ &\leqslant \log \int f(w,\theta)dw = \log 1 = 0 \end{aligned}$$

- Inequality is strict when $\mathbb{P}(f(W_i,\theta) \neq f(W_i,\theta_0)) > 0$ for all $\theta \neq \theta_0 \in \Theta$.
- Hence, $\mathbb{P}(f(W_i,\theta) \neq f(W_i,\theta_0)) > 0$ for all $\theta \neq \theta_0 \in \Theta$ sufficient for ID1.(iii).

Sufficient conditions for ID - ML

- ID1.(iii) [θ_0 uniquely minimises $Q(\theta)$ over Θ] in ML
- Now, true distribution $g(w)$ *not* part of parametric family $\{f(w,\theta) : \theta \in \Theta\}$
- Kullback-Liebler Information Criterion:

$$KLIC(g, f(\cdot,\theta)) = \mathbb{E}_g \log g(W_i) - \mathbb{E}_g \log f(W_i,\theta)$$

$\mathbb{E}_g$ expectations under density g. Note,

$$KLIC(g, f(\cdot,\theta)) = \mathbb{E}_g \log g(W_i) + Q(\theta)$$

- "Quasi-ML": ML under misspecification converges in probability to θ_0 uniquely minimises KLIC between true and paremetised density (provided unique minimiser exists)

Sufficient conditions for ID - LS

- ID1.(iii) [θ_0 uniquely minimises $Q(\theta)$ over Θ] in LS
- First, correctly specified case, i.e. $\theta_0 \in \Theta$ such that $\mathbb{E}[Y_i|X_i] = g(X_i, \theta_0)$ a.s.
- Then θ_0 minimises $Q(\theta)$

Proof.

$$2(Q(\theta) - Q(\theta_0)) = \mathbb{E}(Y_i - g(X_i, \theta) + g(X_i, \theta_0) - g(X_i, \theta_0))^2 - \mathbb{E}(U_i)^2 = \cdots \geqslant 0$$

■

- Last inequality is strict for $\mathbb{P}[g(X_i, \theta) \neq g(X_i, \theta_0)] > 0$ when $\theta \neq \theta_0 \in \Theta$.

Sufficient conditions for ID - LS

- ID1.(iii) [θ_0 uniquely minimises $Q(\theta)$ over Θ] in LS
- Second, incorrectly specified model. Define $h(x) := \mathbb{E}[Y_i|X_i = x]$,

$$\begin{aligned} Q(\theta) &= \mathbb{E}[(Y_i - h(X_i) + h(X_i) - g(X_i, \theta))^2/2] \\ &= \mathbb{E}[(Y_i - h(X_i))^2/2] + \mathbb{E}[h(X_i) - g(X_i, \theta))^2/2] \end{aligned}$$

- LS estimator gives best MSE approximation of $\mathbb{E}[Y_i|X_i = x]$ from family $\{g(\cdot, \theta) : \theta \in \Theta\}$

Sufficient conditions for ID - GMM

- ID1.(iii) [θ_0 uniquely minimises $Q(\theta)$ over Θ] in GMM
- If A nonsingular and exists unique $\theta_0 \in \Theta$ s.t. $\mathbb{E}g(W_i, \theta_0) = 0$, θ_0 uniquely minimises $Q(\theta)$ over Θ

Sufficient conditions for ID - MD

- ID1.(iii) [θ_0 uniquely minimises $Q(\theta)$ over Θ] in MD
- A nonsingular and exists unique $\theta_0 \in \Theta$ s.t. $\pi_0 = g(\theta_0)$, $\theta_0 \in \Theta$ uniquely minimises $Q(\theta)$
- If specifications misspecified, and no $\theta \in \Theta$ s.t. $\pi_0 = g(\theta)$, then,

$$\theta_0 = \underset{\theta \in \Theta}{\operatorname{argmin}}\, Q(\theta) = \underset{\theta \in \Theta}{\operatorname{argmin}}\, \|A(\pi_0 - g(\theta))\|/2$$

if this minimiser exists

Sufficient conditions for ID - TS

- ID1.(iii) [θ_0 uniquely minimises $Q(\theta)$ over Θ] in TS
- A nonsingular, exists unique $\theta_0 \in \Theta$ s.t. $G(\theta_0, \tau_0) = 0$, $\theta_0 \in \Theta$ uniquely minimises $Q(\theta)$

Sufficient conditions for U-WCON

- ULLN is a uniform, hence stronger, assumption than pointwise restrictions
- Would like to relax uniformity, but still imply U-WCON under weak sufficient conditions

Assumption 5 (P-WCON)

$\forall \theta \in \Theta,\ Q_n(\theta) \xrightarrow{p} Q(\theta)$

- Compare to earlier uniform restriction: $\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| = o_p(1)$

Lemma 2 (WLLN)

$\{W_i\}_{i=1}^n$ i.i.d. r.v., $\{m(w,\theta) : \theta \in \Theta\}$ class of $\mathbb{R}^k$-functions with $\mathbb{E}[\|m(W_i,\theta)\|] < \infty\ \forall \theta \in \Theta$. Then, as $n \to \infty$, $\forall \theta \in \Theta$,

$$\frac{1}{n}\sum_{i=1}^{n} m(W_i,\theta) \xrightarrow{p} \mathbb{E}[m(W_i,\theta)]$$

Sufficient conditions for U-WCON

- P-WCON holds for our considered estimators, e.g.:
  1. ML: $\{W_i\}_{i=1}^n$ i.i.d., $\mathbb{E}[|\log f(X_i,\theta)|] < \infty \ \forall\theta \in \Theta$
  2. LS: $\{W_i\}_{i=1}^n$ i.i.d., $\mathbb{E}[(Y_i - g(X_i,\theta))^2] < \infty \ \forall\theta \in \Theta$
  3. MD: $A_n \xrightarrow{p} A$ and $\widehat{\pi}_n \xrightarrow{p} \pi_0$
  4. TS: $A_n \xrightarrow{p} A$, $\widehat{\tau}_n \xrightarrow{p} \tau_0$, $\sup_{\tau\in B(\tau_0,\varepsilon)} |G_n(\theta,\tau) - G(\theta,\tau)| \xrightarrow{p} 0$ for some $\epsilon > 0 \ \forall\theta \in \Theta$ and $G(\theta,\tau)$ continuous at $\tau_0 \ \forall\theta \in \Theta$, from,

$$\begin{aligned}
|G_n(\theta,\widehat{\tau}_n) - G(\theta,\tau)| &\leqslant |G_n(\theta,\widehat{\tau}_n) - G(\theta,\widehat{\tau}_n)| + |G(\theta,\widehat{\tau}_n) - G(\theta,\tau_0)| \\
&\leqslant \sup_{\tau\in B(\tau_0,\varepsilon)} |G_n(\theta,\widehat{\tau}) - G(\theta,\widehat{\tau})| + |G(\theta,\widehat{\tau}_n) - G(\theta,\tau_0)| \\
&\xrightarrow{p} 0.
\end{aligned}$$

Sufficient conditions for U-WCON – Stochastic Equicontinuity

- Define $H_n(\theta) = Q_n(\theta) - Q(\theta)$

Definition 1

$\{H_n(\theta) : n \geqslant 1\}$ *is stochastically equicontinuous on* Θ *if* $\forall\epsilon > 0$, $\exists\delta > 0$ *s.t.*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{\theta\in\Theta}\sup_{\theta'\in B(\theta,\delta)} |H_n(\theta) - H_n(\theta')| > \epsilon\right) < \epsilon$$

- Equicontinuity: continuous and equal variation over given neighbourhoods
- $\mathcal{F}$ equicontinuous at $x_0 \in X$ if $\forall\epsilon > 0$ $\exists\delta = \delta(x_0) > 0$ s.t. $d(f(x_0), f(x)) < \epsilon$ $\forall f \in \mathcal{F}$ and $\{x : d(x_0, x) < \delta\}$. $\mathcal{F}$ equicontinuous if equicontinuous at every $x \in X$.
- Uniform equicontinuity: sequence $\{f_n(x), n = 1, 2, \ldots\}$ equicontinuous at x_0, i.e., $\forall\epsilon > 0$ $\exists\delta = \delta(x_0) > 0$ s.t. $d(f(x_0), f(x)) < \epsilon$ for all n and $\{x : d(x_0, x) < \delta\}$. For uniformity, δ cannot depend on n (but can on x_0). *When* exists δ that applies to all $x \in X$, sequence is uniformly equicontinuous.

Sufficient conditions for U-WCON – Stochastic Equicontinuity

Assumption 6 (SE)

$\{H_n(\theta) : n \geqslant 1\}$ is SE on Θ

Theorem 2

(a) *If Θ bounded, P-WCON and SE hold, then $\sup_{\theta\in\Theta} |H_n(\theta)| \xrightarrow{p} 0$.*

(b) *If $\sup_{\theta\in\Theta} |H_n(\theta)| \xrightarrow{p} 0$, then P-WCON and SE hold.*

Sufficient conditions for SE

Assumption 7 (SE1)

(i) W_i identically distributed; (ii) $m(w, \theta)$ continuous in $\theta \in \Theta$ $\forall w$ in support of $\mathcal{W}$; (iii) $\mathbb{E}[\sup_{\theta\in\Theta} \|m(W_i, \theta)\|] < \infty$; (iv) Θ compact.

Theorem 3

Assumption SE1 holds. Then, SE holds and $\mathbb{E}[\|m(W_i, \theta)\|]$ continuous in θ at θ_0. Further, suppose $\{W_i\}_{i=1}^n$ independent (or stationary and ergodic) dist., then P-WCON, therefore U-WCON hold.

Proof.

$$v_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} \big(m(W_i, \theta) - \mathbb{E}[m(W_i, \theta)]\big)$$

in class... ■

References

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of econometrics 18*(1), 5–46.

Chamberlain, G. (1984). Panel data. *Handbook of Econometrics 2*.