Large Sample Properties of Extremum Estimators

Robert A. Miller

Econometrics 2

November 2019

- Let $h_N(\omega, \theta) : \Omega \times \Theta \rightarrow \mathbb{R}^q$ be a (composite) function of the sample, measurable with respect to $(\Omega, \mathcal{F}_{N,} P)$.
- For $h = (h_1, ..., h_q)$, define $\|h\| = \sqrt{\sum_{i=1}^{q} h_i^2}$.
- Also let $\theta_N(\omega)$ be a random variable with respect to $(\Omega, \mathcal{F}_{N,} P)$ that depends on ω through the data.
- Then $\theta_N(\omega)$ is called an extremum (that is M for maximum or minimum) estimator if and only if

$$\|h_N(\omega, \theta_N(\omega))\| \leq o_p(1) + \inf_{\theta \in \Theta} \|h_N(\omega, \theta)\| \tag{1}$$

- For example in a GMM framework we choose $\theta \in \Theta$ to minimize:

$$\left[\frac{1}{N}\sum_{n=1}^{N} f_N\left(x_N\left(\omega\right),\theta\right)\right]' A_N\left(\omega\right)\left[\frac{1}{N}\sum_{n=1}^{N} f_N\left(x_N\left(\omega\right),\theta\right)\right]$$

- Since A is positive definite, and without loss of generality symmetric, we may:
  - factor $A_N\left(\omega\right)$ into:

$$A_N\left(\omega\right) = B_N\left(\omega\right)' B_N\left(\omega\right) \text{ where } B_N\left(\omega\right) \text{ is } q \times q$$

  - define:

$$h_N\left(\omega,\theta\right) \equiv B'_N\left(\omega\right) N^{-1}\sum_{n=1}^{N} f_N\left(x_N\left(\omega\right),\theta\right)$$

  - and note GMM minimizes $\left\|h_N\left(\omega,\theta\right)\right\|$ with respect to θ.

Theorem

Suppose the DGP is represented by $\theta_0 \in \Theta$. *If:*

$$\|h_N(\omega, \theta_0)\| = o_p(1) \tag{2}$$

and:

$$\sup_{\|\theta - \theta_0\| > \frac{1}{r}} \|h_N(\omega, \theta)\|^{-1} = O_p(1) \tag{3}$$

for all $r \in \{1, 2, \ldots\}$, *then* $\theta_N = \theta_0 + o_p(1)$.

- Intuitively, the criterion function must converge to zero in probability for the true DGP, but if another DGP is used then the inverse of the criterion function should be bounded.
- These sufficient conditions for a consistent estimator mimic necessary and sufficient conditions for identification.

To prove the theorem we start with a preliminary lemma, draw out some implications of convergence in probability and then collect the results.

Lemma

If

$$\sup_{\|\theta-\theta_0\|>\frac{1}{r}} \|h_N(\omega,\theta)\|^{-1} \leq M \tag{4}$$

and

$$\|h_N(\omega,\theta_N(\omega))\|^{-1} > M, \tag{5}$$

then

$$\|\theta_N(\omega)-\theta_0\| \leq \frac{1}{r} \tag{6}$$

A Theorem on Consistency

Proof of preliminary lemma

Proof.

We establish this claim by a contradiction argument:
Suppose (4) and (5) are true, but not (6). Then:

$$\|\theta_N(\omega) - \theta_0\| > \frac{1}{r} \tag{7}$$

Hence:

$$\begin{aligned} M &< \|h_N(\omega, \theta_N(\omega))\|^{-1} && \text{by (5)} \\ &\leq \sup_{\|\theta-\theta_0\|>\frac{1}{r}} \|h_N(\omega, \theta)\|^{-1} && \text{by (7)} \\ &\leq M && \text{by (4).} \end{aligned}$$

This contradiction proves the complement to the lemma is false. □

The preliminary lemma implies:

$$
\begin{aligned}
&\Pr\left(\left\{\omega : \|\theta_N(\omega) - \theta_0\| > \frac{1}{r}\right\}\right) \\
&\leq \Pr\left(\begin{array}{c} \left\{\omega : \|h_N(\omega, \theta_N(\omega))\|^{-1} \leq M\right\} \\ \cup \left\{\omega : M < \sup_{\|\theta - \theta_0\| > \frac{1}{r}} \|h_N(\omega, \theta)\|^{-1}\right\} \end{array}\right) \\
&\leq \begin{array}{c} \Pr\left(\left\{\omega : \|h_N(\omega, \theta_N(\omega))\|^{-1} \leq M\right\}\right) \\ + \Pr\left(\left\{\omega : M < \sup_{\|\theta - \theta_0\| > \frac{1}{r}} \|h_N(\omega, \theta)\|^{-1}\right\}\right) \end{array}
\end{aligned}
$$

Appealing to (3) the definition of $O_p(1)$, for all $\varepsilon > 0$ there exists a real number M_ε such that for each $r \in \{1, 2, \ldots\}$:

$$\Pr\left(\left\{\omega : M_\varepsilon < \sup_{\|\theta - \theta_0\| > \frac{1}{r}} \|h_N(\omega, \theta)\|^{-1}\right\}\right) \leq \varepsilon \tag{8}$$

Also from (2):

$$\|h_N(\omega, \theta_0)\| = o_p(1)$$

It now follows from (1) that:

$$\begin{aligned}
\left\| h_N\left(\omega, \theta_N\left(\omega\right)\right)\right\| &\leq o_p\left(1\right) + \inf_{\theta \in \Theta} \left\| h_N\left(\omega, \theta\right)\right\| \\
&\leq 2o_p\left(1\right) + \left\| h_N\left(\omega, \theta_0\right)\right\| \\
&= 3o_p\left(1\right) = o_p\left(1\right)
\end{aligned}$$

Thus by the definition of convergence in probability, for all $\varepsilon > 0$ and any $r \in \{1, 2, \ldots\}$, there exists N_ε such that for $N > N_\varepsilon$

$$\begin{aligned}
\varepsilon &\geq \Pr\left(\left\{\omega : \left\| h_N\left(\omega, \theta_N\left(\omega\right)\right)\right\| > \frac{1}{r}\right\}\right) \\
&= \Pr\left(\left\{\omega : \left\| h_N\left(\omega, \theta_N\left(\omega\right)\right)\right\|^{-1} < M\right\}\right)
\end{aligned}$$

where $M = r$.

A Theorem on Consistency

Proof of the theorem

Therefore, for $N > N_\varepsilon$

$$
\begin{aligned}
&\Pr\left(\left\{\omega : \|\theta_N(\omega) - \theta_0\| > \frac{1}{r}\right\}\right) \\
&\quad \leq \Pr\left(\left\{\omega : \|h_N(\omega, \theta_N(\omega))\|^{-1} \leq M\right\}\right) \\
&+ \Pr\left(\left\{\omega : M < \sup_{\|\theta - \theta_0\| > \frac{1}{r}} \|h_N(\omega, \theta)\|^{-1}\right\}\right) \\
&\quad \leq \varepsilon + \varepsilon
\end{aligned}
$$

Appealing to the definition of $o_p(1)$ and convergence in probability one more time the theorem is proved.

- First we review the definition of a probability distribution function and describe its properties.
- Then we define convergence in distribution, or weak convergence.
- We show how (un)related weak convergence is to the other forms of convergence we have analyzed in this course.
- This leads us an investigation of the asymptotic distributional properties of extremal or M estimators.
- Focusing on GMM estimators, we analyze the choice of the:
  1. weighting matrix, which balances the orthogonality conditions when there are overidentifying restrictions.
  2. optimal instruments, used in an IV context.
  3. orthogonality conditions, to achieve the asymptotic lower bound.

Convergence in Distribution

Probability distribution functions

- Say x is a random variable with respect to $(\Omega, \mathcal{F}, P)$.
- The (cumulative) probability distribution function for x is defined $\forall b \in \mathbb{R}$ as:

$$F_x(b) \equiv \Pr\{\omega : x(\omega) \leq b\} = \Pr\{x \leq b\}.$$

1. $F_x(\cdot)$ is increasing and right continuous.
2. Also $\lim_{b \to -\infty} F_x(b) = 0$ and $\lim_{b \to \infty} F_x(b) = 1$.
3. Lastly $F_x(b)$ is continuous at b iff $F_x(b - \varepsilon) \to F_x(b)$ as $\varepsilon \to 0$.

Convergence in Distribution

Continuity of distribution function

Lemma

$F_x(\cdot)$ is continuous at b if and only if $\Pr\{\omega : x(\omega) = b\} = 0$.

Proof.

$$
\begin{aligned}
\{\omega : x(\omega) \le b\} &= \{\omega : x(\omega) < b\} \cup \{\omega : x(\omega) = b\} \\
\Rightarrow \quad \Pr\{\omega : x(\omega) \le b\} &= \Pr\{\omega : x(\omega) < b\} + \Pr\{\omega : x(\omega) = b\} \\
\Leftrightarrow \quad \Pr\{\omega : x(\omega) = b\} &= \Pr\{\omega : x(\omega) \le b\} - \Pr\{\omega : x(\omega) < b\} \\
&= F_x(b) - F_x(b^-)
\end{aligned}
$$

RHS $= 0$ if and only if F is continuous. □

- Thus $F_x(\cdot)$ is continuous at b if and only if b has measure zero.

Convergence in Distribution

Convergence in distribution defined

- Let $F_1, F_2, \ldots$ be distribution functions for $x_1, x_2, \ldots$
- We say x_N converges weakly (in distribution or in law) to x, that is $x_N \xrightarrow{d} x$, if and only if:

$$F_N(b) \rightarrow F(b) \quad \text{for all } b \text{ where } F \text{ is continuous.}$$

- For example let:

$$F_N(b) = \begin{cases} 0 & b < \frac{1}{N} \\ 1 & b \geq \frac{1}{N} \end{cases}$$

$$F(b) = \begin{cases} 0 & b < 0 \\ 1 & b \geq 0 \end{cases}$$

- In contrast with the other convergence concepts, weak convergence is not related the mapping $x_N(\omega) : \Omega \rightarrow \mathbb{R}$.

Convergence in Distribution

Relating convergence in probability to convergence in distribution

Lemma

If $x_N \xrightarrow{p} x$, *then* $x_N \xrightarrow{d} x$.

Proof.

See Dhrymes *Topics in Advanced Econometrics* (1990 pages 161-162). □

Lemma

$x_N \xrightarrow{d} x$ *does not imply* $x_N \xrightarrow{p} x$.

Proof.

(by construction)

Let $x \sim \mathcal{N}(0,1)$ and $x_N = (-1)^{N+1} x$.

Then (trivially) $x_N \xrightarrow{d} x$ but $x_N \overset{p}{\nrightarrow} x$. □

Convergence in Distribution

Weak convergence to the same distribution

Lemma

Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be two sequences of random variables with respect to $(\Omega, \mathcal{F}, P)$:

$$\text{If } X_N - Y_N = o_p(1) \text{ and } X_N \xrightarrow{d} X \text{ then } Y_N \xrightarrow{d} X$$

Proof.

See Fuller (1976, Theorem 5.2.1. pages 193 -194): "Introduction to Statistical Time Series" □

- To analyze the stochastic properties of GMM estimators, we introduce notation to make explicit their dependence on the history created by revealing an ever-increasing data set.
- We denote:
  - the path that would be taken by a complete countable data set by the element $\omega \in \Omega$, one of many paths that could have been taken.
  - the increasing sequence of σ-algebras, denoted $\sigma_1 \subseteq \sigma_2 \subseteq \ldots$, to represent the information the data with data sets of size $N = 1, 2, \ldots$.
  - a probability measure P associated with the measure space (Ω, σ, P)
  - $x_n(\omega)$ a vector of random variables measurable with respect to σ_n. the information in a data set by random variables

- Armed with this expanded view of the data, we recall the "normal equations" definition of a GMM estimator for $\theta_0 \in \Theta$ as

$$0 = A_N^* (\omega) \frac{1}{N} \sum_{n=1}^{N} f (x_n (\omega), \theta_N (\omega))$$

which is the FOC for minimizing:

$$\left[\frac{1}{N} \sum_{n=1}^{N} f (x_n (\omega), \theta) \right]' A_N (\omega) \left[\frac{1}{N} \sum_{n=1}^{N} f (x_n (\omega), \theta) \right]$$

with respect to θ where by the mean value theorem, for some $\theta_N^* (\omega) \xrightarrow{p} \theta_N (\omega)$:

$$A_N^* (\omega) = \left(\frac{1}{N} \sum_{n=1}^{N} \frac{\partial f (x_n (\omega), \theta_N^* (\omega))}{\partial \theta} \right)' A_N (\omega)$$

Asymptotic Distribution of GMM estimators

More notation

- Denote by:

$$
\begin{aligned}
D_N(\omega) &\equiv \frac{1}{N}\sum_{n=1}^{N}\frac{\partial f\left(x_n(\omega),\theta_N^*(\omega)\right)}{\partial\theta} \\
D_0 &\equiv E\left[\frac{\partial f}{\partial\theta}\left(x_n(\omega),\theta_0\right)\right] \\
\Sigma_N(\omega) &\equiv \frac{1}{N}\sum_{n=1}^{N}\left[f\left(x_n(\omega),\theta_0\right)f\left(x_{n-j}(\omega),\theta_0\right)'\right] \\
\Sigma_0 &\equiv \sum_{j=-\infty}^{\infty}E\left[f\left(x_n(\omega),\theta_0\right)f\left(x_{n-j}(\omega),\theta_0\right)'\right]
\end{aligned}
$$

Asymptotic Distribution of GMM estimators

Some assumptions

- We assume that for some appropriately defined A_0, D_0 and Σ_0, a law of large numbers guarantees:

$$\theta_N(\omega) \xrightarrow{p} \theta_0 \qquad A_N(\omega) \xrightarrow{p} A_0 \qquad D_N(\omega) \xrightarrow{p} D_0 \qquad \Sigma_N(\omega) \xrightarrow{p} \Sigma_0$$

- We also assume a central limit theorem ensures:

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} f\left(x_n(\omega), \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_0\right)$$

Asymptotic Distribution

Taylor series expansion

- Subsuming the dependence on $\omega \in \Omega$ the FOC or normal equations can be expressed:

$$0 = A_N^* \frac{1}{N} \sum_{n=1}^{N} f\left(x_n, \theta_N\right)$$

- It follows from the mean value theorem that there exists some $\widetilde{\theta}_N$ on the linear interval connecting θ_0 to θ_N such that:

$$\begin{aligned} LHS &\equiv -A_N^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f\left(x_n, \theta_0\right) \\ &= A_N^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \frac{\partial f\left(x_n, \widetilde{\theta}_N\right)}{\partial \theta} \left(\theta_N - \theta_0\right) \equiv RHS \end{aligned}$$

Asymptotic Distribution

Result and outline of proof

- We show below that:

$$LHS = A_0^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f(x_n, \theta_0) + o_p(1)$$

$$RHS = [A_0^* D_0 + o_p(1)] \sqrt{N} (\theta_N - \theta_0)$$

- We then prove equating the *LHS* to the *RHS* implies:

$$A_0^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f(x_n, \theta_0) = A_0^* D_0 \sqrt{N} (\theta_N - \theta_0) + o_p(1)$$

- Since random variables differing by $o_p(1)$ have the same asymptotic distribution:

$$A_0^* D_0 \sqrt{N} (\theta_N - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, A_0^* \Sigma_0 A_0^{*\prime}\right)$$

and hence:

$$\sqrt{N} (\theta_N - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, (A_0^* D_0)^{-1} A_0^* \Sigma_0 A_0^{*\prime} (A_0^* D_0)^{-1\prime}\right)$$

Asymptotic Distribution
The LHS

$$
\begin{aligned}
LHS &\equiv A_N^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f\left(x_n, \theta_0\right) \\
&= A_0^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f\left(x_n, \theta_0\right) + \left(A_N^* - A_0^*\right) \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f\left(x_n, \theta_0\right) \\
&= A_0^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f\left(x_n, \theta_0\right) + o_p\left(1\right) \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f\left(x_n, \theta_0\right) \\
&= A_0^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f\left(x_n, \theta_0\right) + o_p\left(1\right) O_p\left(1\right) \\
&= A_0^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f\left(x_n, \theta_0\right) + o_p\left(1\right)
\end{aligned}
$$

Asymptotic Distribution

The RHS

$$\begin{aligned} RHS &\equiv A_N^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \frac{\partial f\left(x_n, \widetilde{\theta}_N\right)}{\partial \theta} (\theta_N - \theta_0) \\ &= \left[A_0^* + o_p(1)\right]\left[D_0 + o_p(1)\right] \sqrt{N} (\theta_N - \theta_0) \\ &= \left[A_0^* D_0 + o_p(1)\right] \sqrt{N} (\theta_N - \theta_0) \end{aligned}$$

since:

$$\begin{aligned} & \left[A_0^* + o_p(1)\right]\left[D_0 + o_p(1)\right] \\ = \; & A_0^* D_0 + o_p(1) A_0^* + o_p(1) D_0 + o_p^2(1) \\ = \; & A_0^* D_0 + o_p(1) \end{aligned}$$

Asymptotic Distribution

Equating the LHS and the RHS

- Equating the expressions for the *LHS* and the *RHS* we obtain:

$$A_0^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f(x_n, \theta_0) + o_p(1) = [A_0^* D_0 + o_p(1)] \sqrt{N} (\theta_N - \theta_0)$$

- Noting $A_0^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f(x_n, \theta_0)$ is $O_p(1)$ it follows by a simple contradiction argument that $\sqrt{N}(\theta_N - \theta_0)$ is also $O_p(1)$, so:

$$\begin{aligned} & A_0^* \frac{1}{\sqrt{N}} \sum_{n=1}^{N} f(x_n, \theta_0) \\ = \;& [A_0^* D_0 + o_p(1)] \sqrt{N} (\theta_N - \theta_0) + o_p(1) \\ = \;& A_0^* D_0 \sqrt{N} (\theta_N - \theta_0) + o_p(1) \sqrt{N} (\theta_N - \theta_0) + o_p(1) \\ = \;& A_0^* D_0 \sqrt{N} (\theta_N - \theta_0) + o_p(1) O_p(1) + o_p(1) \\ = \;& A_0^* D_0 \sqrt{N} (\theta_N - \theta_0) + o_p(1) \end{aligned}$$