Large Sample Properties of Extremum Estimators

Robert A. Miller

Structural Econometrics

November 2021

- Let $h_N(\omega, \theta) : \Omega \times \Theta \to \mathbb{R}^q$ 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 iff:

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

- 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 $A_N(\omega)$ into:

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

  - define:

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

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

Theorem

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

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

and:

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

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 reflect conditions that are necessary and sufficient 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 for any $M \in (0, \infty)$:

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

and:

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

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:

$$
\begin{aligned}
&\Pr\left(\left\{\omega : \left\|\theta_N(\omega) - \theta_0\right\| > \frac{1}{r}\right\}\right) \\
&\leq \Pr\left(\begin{array}{c} \left\{\omega : \left\|h_N\left(\omega, \theta_N(\omega)\right)\right\|^{-1} < M\right\} \\ \cup \left\{\omega : M \leq \sup_{\|\theta-\theta_0\|>\frac{1}{r}} \left\|h_N(\omega, \theta)\right\|^{-1}\right\} \end{array}\right) \\
&\leq \begin{array}{l} \Pr\left(\left\{\omega : \left\|h_N\left(\omega, \theta_N(\omega)\right)\right\|^{-1} \leq M\right\}\right) \\ + \Pr\left(\left\{\omega : M < \sup_{\|\theta-\theta_0\|>\frac{1}{r}} \left\|h_N(\omega, \theta)\right\|^{-1}\right\}\right) \end{array}
\end{aligned} \tag{8}
$$

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

$$\Pr\left(\left\{\omega : M_\varepsilon < \sup_{\|\theta-\theta_0\|>\frac{1}{r}} \|h_N(\omega, \theta)\|^{-1}\right\}\right) \leq \varepsilon \tag{9}$$

Also from (2):

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

A Theorem on Consistency

Proof of theorem (3)

It now follows from (1), the inequalities:

$$0 \leq \inf_{\theta \in \Theta} \|h_N(\omega, \theta)\| \leq \|h_N(\omega, \theta_0)\| = o_p(1)$$

and (2) that:

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

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

$$\begin{aligned} \varepsilon &\geq \Pr\left(\left\{\omega : \|h_N(\omega, \theta_N(\omega))\| > \frac{1}{r}\right\}\right) \\ &= \Pr\left(\left\{\omega : \|h_N(\omega, \theta_N(\omega))\|^{-1} < M\right\}\right) \end{aligned} \quad (10)$$

A Theorem on Consistency

Proof of theorem (4)

Therefore from (8), for $N > N_{\varepsilon,r}$:

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

because:

- $\Pr\left(\left\{\omega : \|h_N(\omega, \theta_N(\omega))\|^{-1} < M\right\}\right) \leq \varepsilon$ by (10).
- and $\Pr\left(\left\{\omega : M < \sup_{\|\theta-\theta_0\|>\frac{1}{r}} \|h_N(\omega, \theta)\|^{-1}\right\}\right) \leq \varepsilon$ by (9).

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

- We now turn to 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.

Asymptotic Distribution of GMM estimators

The underlying probability space

- 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 by:
    - $\omega \in \Omega$, a path that a complete history could take.
    - $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$, an increasing sequence of σ-algebras, to represent potential for gathering information with $N = 1, 2, \ldots$.
    - P, a probability measure associated with the measure space $(\Omega, \mathcal{F}, P)$.
    - $X_n(\omega)$, a vector of random variables measurable with respect to $\mathcal{F}_n$, showing what is gathered.
    - $\widetilde{\omega} \in \Omega$, the particular draw of history upon which all data is based.
    - $X_n(\widetilde{\omega})$, the n^{th} observation of the unfolding history $\widetilde{\omega}$.
    - $\theta_0 \in \Theta$, the parameter determining the DGP.

Asymptotic Distribution of GMM estimators

Defining GMM estimators as the solution to a set of equations

- Armed with this new interpretation of the data, recall the "normal equations" definition of a GMM estimator for θ_0 is found by solving:

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

where $A_N^* \left(\widetilde{\omega}\right)$ converges in probability to:

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

which is the FOC for minimizing:

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

with respect to θ.

Asymptotic Distribution of GMM estimators

More notation

- Denote by:

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

- In constructing $\Sigma_N(\widetilde{\omega})$, note that the $n-j$ sequence is $N, N-1, \ldots, 1$ whereas it is $\ldots n+1, n, n-1, \ldots$ in Σ_0, population analogue.

Asymptotic Distribution of GMM estimators

Some assumptions

- Given $\theta_0 \in \Theta$ and $f(x, \theta) : \mathbb{R}^p \times \Theta \to \mathbb{R}^q$, we assume that a LLN guarantees:

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

  for some appropriately defined $q \times q$ matrix A_0.

- We also assume a CLT ensures:

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

- To simplify notation, we subsume sample dependence on $\widetilde{\omega}$ and write:

$$x_n \equiv X_n(\widetilde{\omega}) \qquad A_N \equiv A_N(\widetilde{\omega}) \qquad D_N \equiv D_N(\widetilde{\omega}) \qquad A_N^* = A_N^*(\widetilde{\omega})$$

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

Asymptotic Distribution

Main result

- The assumptions above imply:

$$A_N^* \xrightarrow{p} D_0' A_0 \equiv A_0^*$$

and:

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

- Note:

$$\Psi_0 \equiv \left(A_0^* D_0\right)^{-1} A_0^* \Sigma_0 A_0^{*\prime} \left(A_0^* D_0\right)^{-1\prime}$$

which can be estimated with:

$$\Psi_N \equiv \left[A_N^* D_N\right]^{-1} A_N^* \Sigma_N A_N^{*\prime} \left[A_N^* D_N\right]^{-1\prime}$$

- For example if A_N is the identity matrix, then:

$$A_N^* = D_N' A_N = D_N'$$

and Ψ_N simplifies to:

$$\left[D_N' D_N\right]^{-1} D_N' \Sigma_N D_N \left[D_N' D_N\right]^{-1\prime}$$

- Another alternative is to set $A_N = \Sigma_N^{-1}$, or more generally a consistent estimator of Σ_N. Then:
  1. $A_N \xrightarrow{p} \Sigma_0^{-1}$.
  2. $A_N^* \xrightarrow{p} D_0' \Sigma_0^{-1} = A_0^*$.
  3. $A_N^* \Sigma_N A_N^* \xrightarrow{p} D_0' \Sigma_0^{-1} \Sigma_0 \Sigma_0^{-1} D_0 = D_0' \Sigma_0^{-1} D_0$.
  4. $\Rightarrow \Psi \xrightarrow{p} (A_0^* D_0)^{-1} D_0' \Sigma_0^{-1} D_0 (A_0^* D_0)^{-1\prime} = \left(D_0' \Sigma_0^{-1} D_0\right)^{-1}$.

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:

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

Asymptotic Distribution

Outline of proof

- We show below that:

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

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

- Then we prove that 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

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

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

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:

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