

Review fixed effects on your own time!

Agenda

- Extremum Estimation
- GMM vs Extremum Estimation
- Examples
- M- estimation
- Consistency Thm ①
 - Motivation
 - Hansen proof
 - In class proof
- Consistency Thm ②
- Consistency Thm ③

Extremum Estimation

An extremum estimator is any estimator defined as

$$\hat{\theta} = \operatorname{arg\,min} Q_n(w_1, \dots, w_n; \theta)$$

for some parameter θ in parameter space Θ
and w_1, \dots, w_n is the sample

- $Q_n(\cdot) =$ criterion/objective function
 - ↳ indexed by n because it must depend on your sample
 - ↳ $Q(\cdot) =$ population analog of $Q_n(\cdot)$

Note: Under extremum estimation, we make no assumptions about $Q_n(\cdot)$ aside from the fact that $Q_n(\cdot)$ must be a function of the data (w_1, \dots, w_n) and θ .

- $Q_n(\cdot)$ does not necessarily need to be in some quadratic form (like GMM)
 - ↳ encompasses nonlinear objective functions that may require calculation through numerical methods
 - ↳ ie: may not have closed form solution

Aside from the assumptions later on, Extremum Estimation only requires that we minimize/maximize an objective function.

Examples of Extremum Estimation

- Nonlinear Least Squares

$$\begin{aligned} Q(\theta) &= \mathbb{E}(Y - m(X, \theta))^2 \\ Q_n(\theta) &= \mathbb{E}_n(Y - m(X, \theta))^2. \end{aligned}$$

- GMM

$$\begin{aligned} Q(\theta) &= \mathbb{E}g(\theta)' \mathbf{W} \mathbb{E}g(\theta) \\ Q_n(\theta) &= \mathbb{E}_n g(\theta)' \hat{\mathbf{W}} \mathbb{E}_n g(\theta). \end{aligned}$$

- Maximum Likelihood Estimation

$$\begin{aligned} Q(\theta) &= \mathbb{E}\ell(W; \theta) \\ Q_n(\theta) &= \mathbb{E}_n \ell(W; \theta). \end{aligned}$$

M-Estimation

Special case of Extremum Estimation where your criterion function is defined as:

$$\begin{aligned} Q(\theta) &= \mathbb{E}m(W; \theta) \\ Q_n(\theta) &= \mathbb{E}_n m(W; \theta) \end{aligned}$$

where $m(\cdot)$ is some known real-valued function

Note: M-estimation only requires that your criterion function be in the form of a sample avg.

Ex: If we take the log of the likelihood, we end up solving for

$$\begin{aligned} \hat{\theta} &= \arg \max_{\theta \in \Theta} \mathbb{E}_n m(W, \theta) \\ &= \arg \max_{\theta \in \Theta} \underbrace{\frac{1}{n} \sum \log f(Y_i | X_i, \theta)}_{\text{log-likelihood}} \end{aligned}$$

Consistency

Note: Since there are multiple versions of consistency theorems for extremum estimation, I will give them unofficial names

[Consistency Thm A]

Theorem

Assume:

- The sample criterion uniformly consistently estimates the population criterion:

$$\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow{P} 0.$$

- θ_0 is a unique and well-separated global minimum of $Q(\cdot)$:

$$\forall \epsilon > 0 \exists \delta > 0 : Q^\epsilon \equiv \inf_{\theta \in \Theta : \|\theta - \theta_0\| \geq \epsilon} Q(\theta) \geq Q(\theta_0) + \delta.$$

Then $\hat{\theta} \xrightarrow{P} \theta_0$.

To illustrate why we need these assumptions to conclude consistency, we will

- setup the motivation for these assumptions
- define the terms
- go back and see why these assumptions fix our issues

[Consistency Thm A] : Uniform Convergence

(Hausen p 781)

(ex: $\hat{\beta} \xrightarrow{P} \beta$)

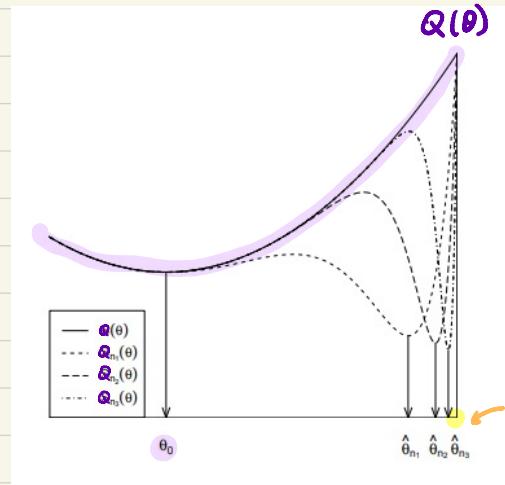
In linear estimators, we apply WLLN to conclude consistency

Issue: Nonlinear estimators may not have a closed form solution
 ↳ cannot easily conclude consistency using WLLN

However, we have the sample criterion function $Q_n(\theta)$,
 which is a sample average

⇒ Hence, $Q_n(\theta) \xrightarrow{P} Q(\theta)$ by WLLN

BUT, if $Q_n(\theta) \xrightarrow{P} Q(\theta)$ $\stackrel{?}{\Rightarrow} \hat{\theta} \xrightarrow{P} \theta ?$



- True function $Q(\theta)$ is a curve
- However, we can find a $Q_n(\theta)$ where:
 - $Q_n(\theta) \rightarrow Q(\theta)$
 - $Q_n(\theta)$ has a sharp dip at the limit of the parameter space

$\hat{\theta}$ in the limit

Hence, we have:

$Q_n(\theta) \xrightarrow{P} Q(\theta)$ but $\hat{\theta}_n \not\xrightarrow{P} \theta_0$

* Uniform Convergence

Definition 22.1 $S_n(\theta)$ converges in probability to $S(\theta)$ uniformly over $\theta \in \Theta$ if

$$\sup_{\theta \in \Theta} |S_n(\theta) - S(\theta)| \xrightarrow{p} 0$$

as $n \rightarrow \infty$.

- convergence happens uniformly across the entire parameter space
- can't have regions of θ where convergence is slow

In other words, for n sufficiently large, the sample criterion function satisfies

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P \left(\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| < \varepsilon \right) = 1$$

- For large n , behavior of $Q_n(\theta)$ is bounded around $[Q(\theta) - \varepsilon, Q(\theta) + \varepsilon]$

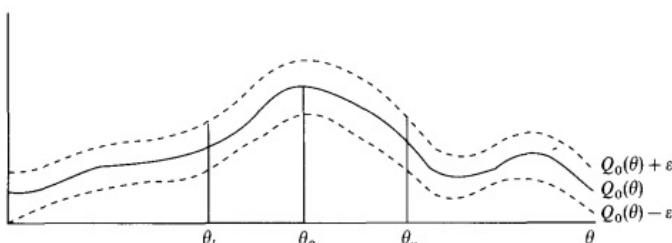


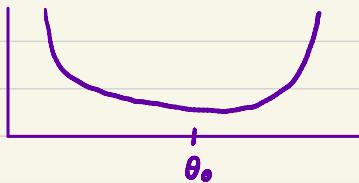
Figure 2.

* Well-Separated

Obviously, our true θ_0 must be unique.

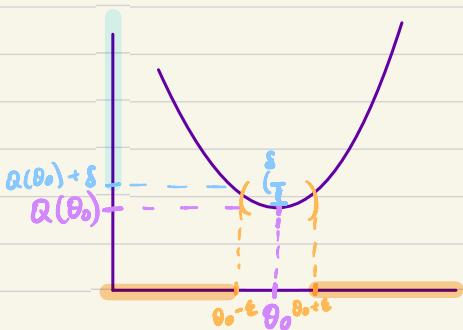
However, if we have multiple θ -values near θ_0 that almost minimize $Q_n(\theta)$, then we might have an issue with slow convergence

Ex:



Definition: θ_0 is well-separated if

$$\forall \epsilon > 0 \exists \delta > 0 : Q^\epsilon \equiv \inf_{\theta \in \Theta : \|\theta - \theta_0\| \geq \epsilon} Q(\theta) \geq Q(\theta_0) + \delta.$$



i.e:

(not in $B_\epsilon(\theta_0)$)
Choose $\theta \in \Theta, \|\theta - \theta_0\| \geq \epsilon$ that make $Q(\theta)$ as small as possible, then my $Q(\theta)$ is still larger than $Q(\theta_0) + \delta$

Consistency Thm A : Hansen's Version

Theorem 22.1 $\hat{\theta} \xrightarrow{p} \theta_0$ as $n \rightarrow \infty$ if

1. $S_n(\theta)$ converges in probability to $S(\theta)$ uniformly over $\theta \in \Theta$.
2. θ_0 uniquely minimizes $S(\theta)$ in the sense that for all $\epsilon > 0$,

$$\inf_{\theta: \|\theta - \theta_0\| \geq \epsilon} S(\theta) > S(\theta_0).$$

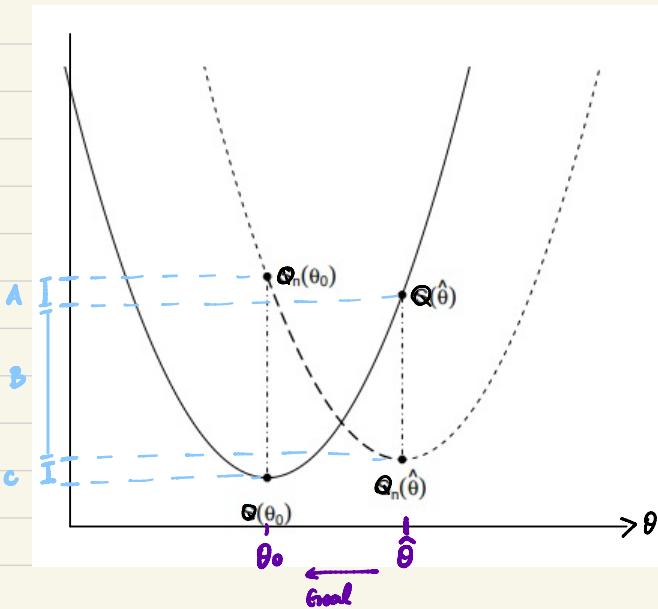
Goal: Prove $\hat{\theta} \xrightarrow{P} \theta_0$

Issue: $Q_n \rightarrow Q$ in order for
sample mm \rightarrow true mm

Proof Structure:

1) $Q(\hat{\theta}) \xrightarrow{P} Q(\theta_0)$

2) $\hat{\theta} \xrightarrow{P} \theta_0$



Consistency Thm A : Hansen's Version

Key Points to Remember:

- (Population) θ_0 minimizes $Q(\theta)$
- (Sample) $\hat{\theta}$ minimizes $Q_n(\theta)$ \star
- θ_0 may not minimize $\underbrace{Q_n(\theta)}_{\text{sample}}$

Proof i) Show $Q(\hat{\theta}) \xrightarrow{P} Q(\theta_0)$

Since θ_0 minimizes $Q(\theta)$ then $Q(\theta_0) \leq Q(\theta)$

$$\Rightarrow 0 \leq Q(\hat{\theta}) - Q(\theta_0)$$

$$= Q(\hat{\theta}) - Q_n(\hat{\theta}) + Q_n(\theta_0) - Q(\theta_0) + Q_n(\hat{\theta}) - Q_n(\theta_0)$$

Hence, $Q(\hat{\theta}) \xrightarrow{P} Q(\theta)$

2) Show $\hat{\theta} \xrightarrow{P} \theta_0$

Let $\varepsilon > 0$ and assume we have a unique minimum

Then $\exists \delta > 0$ s.t

$$\|\theta_0 - \theta\| > \varepsilon \Rightarrow Q(\theta) - Q(\theta_0) \geq \delta$$

Since the above statement is for any $\theta \in \Theta$, then

$$\|\theta_0 - \hat{\theta}\| > \varepsilon \Rightarrow Q(\hat{\theta}) - Q(\theta_0) \geq \delta$$

★ $\Rightarrow P(\|\theta_0 - \hat{\theta}\| > \varepsilon) \leq P(Q(\hat{\theta}) - Q(\theta_0) \geq \delta)$

Hence,

$$P(\|\theta_0 - \hat{\theta}\| > \varepsilon) \xrightarrow{P} 0$$

$$\Rightarrow \hat{\theta} \xrightarrow{P} \theta_0$$

Note: If $A \Rightarrow B$, then $P(A) \leq P(B)$.

$A \Rightarrow B$ implies $A \cap B = A$.

Thus by partition theorem,

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap A^c) \\ &= P(A) + P(B \cap A^c) \geq P(A) \\ &\geq 0 \end{aligned}$$

Proof of Consistency Thm A

(In class theorem)

Theorem

Assume:

- ① The sample criterion uniformly consistently estimates the population criterion:

$$\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow{P} 0.$$

- ② θ_0 is a unique and well-separated global minimum of $Q(\cdot)$:

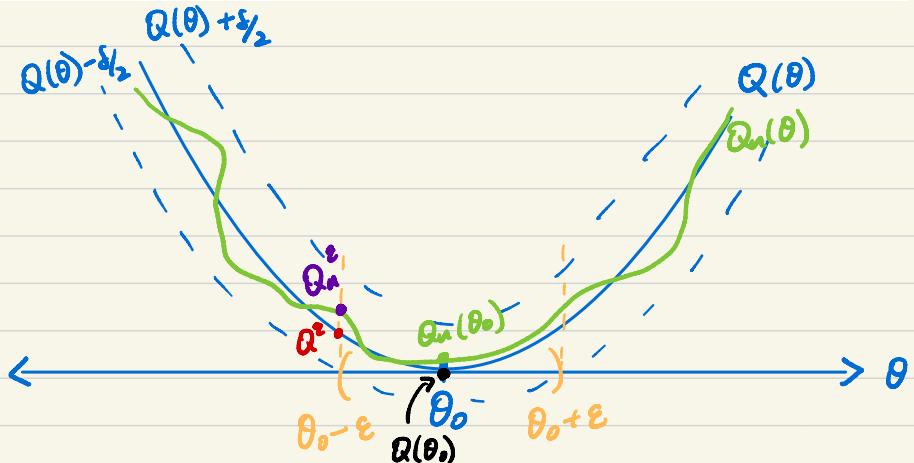
$$\forall \epsilon > 0 \exists \delta > 0 : Q^\epsilon \equiv \inf_{\theta \in \Theta : \|\theta - \theta_0\| \geq \epsilon} Q(\theta) \geq Q(\theta_0) + \delta.$$

Then $\hat{\theta} \xrightarrow{P} \theta_0$.

Proof

Fix $\epsilon > 0$ and define $Q_n^\epsilon \equiv \inf_{\theta \in \Theta : \|\theta - \theta_0\| \geq \epsilon} Q_n(\theta)$, then

$$\begin{aligned} & \Pr(\|\hat{\theta} - \theta_0\| > \epsilon) \\ & \leq \Pr(Q_n^\epsilon \leq Q_n(\theta_0)) \\ & = 1 - \Pr(Q_n^\epsilon > Q_n(\theta_0)) \\ & \leq 1 - \Pr(Q_n^\epsilon > Q_\epsilon - \delta/2, Q_n(\theta_0) < Q(\theta_0) + \delta/2) \\ & \rightarrow 0, \end{aligned}$$



(Prep) Consistency Thm B

Issue: Some of these conditions are hard to check in real life?

Can I create lower-level assumptions such that I can guarantee a well-separated minimum?

Theorem

Assume that:

- ① $Q(\cdot)$ is continuous,
- ② Θ is compact,
- ③ θ_0 uniquely minimizes $Q(\theta)$.

Then θ_0 is a well-separated minimum.

And for uniform convergence:

Theorem

Assume that:

- ① $\hat{\theta}$ is an m-estimator,
- ② The data are i.i.d. realizations of W ,
- ③ $m(W; \theta)$ is a.s. continuous in θ ,
- ④ $|m(W; \theta)| \leq G(W)$ for some function G s.t. $\mathbb{E}G(W) < \infty$,
- ⑤ Θ is compact.

Then $Q_n(\cdot)$ converges to $Q(\cdot)$ uniformly.

Proof

This is the Uniform Law of Large Numbers.

Recall:

$$\begin{aligned} Q(\theta) &= \mathbb{E}m(W; \theta) \\ Q_n(\theta) &= \mathbb{E}_n m(W; \theta) \end{aligned}$$

Consistency Thm (B)

Consolidated Theorem

Assume that:

- ① $Q(\cdot)$ is continuous,
- ② Θ is compact,
- ③ θ_0 uniquely minimizes $Q(\theta)$,
- ④ $\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow{P} 0$.

Then $\hat{\theta} \xrightarrow{P} \theta_0$.

Consistency Thm C

Consistency for Convex $Q(\cdot)$

Assume that:

- ① Θ is convex,
- ② $\theta_0 \in \text{int } \Theta$,
- ③ θ_0 uniquely minimizes $Q(\theta)$,
- ④ $Q_n(\cdot)$ is convex,
- ⑤ $|Q_n(\theta) - Q(\theta)| \xrightarrow{P} 0, \forall \theta \in \Theta$.

Then $\hat{\theta} \xrightarrow{P} \theta_0$.