Lecture 10: Asymptotic Theory of M-Estimators

Econometrics 2 — *Consistency via Uniform Convergence & Asymptotic Identification*

Instructor: Prof. S. Arvanitis | **Notes & digitisation:** T. Kourtalis

Semester: Spring 2026

▷ **Amber = Handwritten Notes (professor's words)**

◇ **Teal = Student's Notes (added explanations)**

Source Material for This Lecture

This document combines two sources:

1. **Handwritten lecture notes** (amber boxes): the professor's in-class presentation of the general asymptotic theory for M-estimators.
2. **Official LAD slides** (26 pages, available on eClass): these slides are *not* reproduced here. They are referenced throughout by page number (e.g. [Slides p. 20]) and should be read alongside this document.

How the two sources relate: The handwritten notes provide the *theorem*. The LAD slides provide the *empirical demonstration* via Monte Carlo. Specifically:

- The slides' **generic design** [Slides p. 13] is the case where the theorem's conditions hold ⇒ the MC confirms consistency [Slides p. 20], [Slides p. 22]–[Slides p. 23].
- The slides' **degenerate design** [Slides p. 14]–[Slides p. 15] is the case where the theorem's conditions *fail* ⇒ the MC shows failure [Slides p. 21], [Slides p. 24].

The student notes (teal boxes) add geometric intuition, figures, and explicit cross-references between the theorem and the MC results.

1 What Problem Does This Lecture Solve?

▷ **Handwritten Notes** (what the professor said)

Reminder: In the framework of a well-specified (semi-)parametric statistical model, we have an observable objective function $\mu_n(\beta)$ and the corresponding M-estimator:

$$\hat{\beta}_n \in \arg\min_{\beta\in\Theta} \mu_n(\beta).$$

Question: Are there sufficient conditions for the weak consistency of $\hat{\beta}_n$?

◇ **Student's Notes**

Why this question matters now:
In Lecture 9 the LAD slides showed two Monte Carlo experiments with dramatically different outcomes [Slides p. 20]–[Slides p. 21]. The generic design showed clean convergence; the degenerate design showed persistent estimation error. The slides did not provide a *theorem* explaining why. This lecture provides that theorem.

Why we cannot just study $\hat{\beta}_n$ directly:
For OLS we had a closed form $\hat{\beta}_n = (X'X)^{-1}X'Y$ (Lecture 3). For LAD, GMM, or MLE with nonlinear moments, there is *no formula* for $\hat{\beta}_n$. The LP reformulation [Slides p. 3]–[Slides p. 5] gives an algorithm to compute $\hat{\beta}_n$, but no algebraic expression to analyse. The two-step route replaces the intractable question "what does $\hat{\beta}_n$ converge to?" with the tractable question "what does $\mu_n(\beta)$ converge to?"

Road-map:

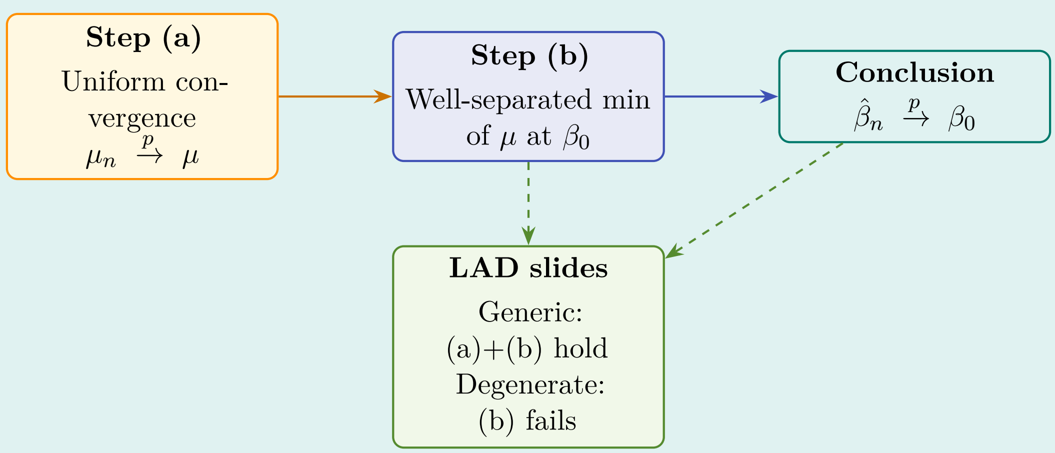

Figure 1: Structure of this lecture. Steps (a) and (b) form the general theorem. The LAD slides (Lecture 9) illustrate both cases: conditions hold (generic) and conditions fail (degenerate).

2 The Main Obstacle and Methodology

▷ **Handwritten Notes** (what the professor said)

Obstacle: Generally we do not know the exact form of the estimator as a function of the sample. One way to work around this is to examine the asymptotic behaviour of the objective function itself as $n \to +\infty$.
The methodology consists of two steps:

(a) Achieving "appropriate convergence" in probability of μ_n to a limit function μ (which depends on β).

(b) If μ satisfies an "asymptotic identification" condition, this implies:

$$\hat{\beta}_n = \arg\min_{\beta \in \Theta} \mu_n(\beta) \xrightarrow{P} \arg\min_{\beta \in \Theta} \mu(\beta) = \beta_0.$$

◇ **Student's Notes**

This is the master theorem of the course.
Every consistency proof we encounter — OLS (Lecture 11), IV (Lecture 12), LAD (Lecture 13) — is a *special case* of verifying conditions (a) and (b) for a specific estimator:

Estimator	How (a) is verified	How (b) is verified
OLS (Lect. 11)	LLN for $X'X/n$, $X'\varepsilon/n$	$Q_{X'X} \succ 0$
IV (Lect. 12)	LLN for $Z'X/n$, $Z'\varepsilon/n$	$Q'_{Z'X} W Q_{Z'X} \succ 0$
LAD generic (Lect. 13)	LLN for $M_n(\beta)$	$M(\beta)$ strictly convex
LAD degenerate [Slides p. 14]	LLN still holds	Flat region: **fails**

3 Step (a): Locally Uniform Convergence in Probability

▷ **Handwritten Notes** (what the professor said)

One form of "appropriate convergence" (though not the only one) is **locally uniform convergence in probability** of μ_n to μ:
For every β there exists a neighbourhood O_β such that:

$$\sup_{\beta^* \in O_\beta} |\mu_n(\beta^*) - \mu(\beta^*)| \xrightarrow{P} 0.$$

◇ **Student's Notes**

Why pointwise convergence is not sufficient:
The LLN directly gives *pointwise* convergence: for each fixed β, $\mu_n(\beta) \xrightarrow{p} \mu(\beta)$. But pointwise convergence does not prevent μ_n from wiggling rapidly near β_0 in a way that shifts the minimiser far from β_0.
Locally uniform convergence says the function μ_n settles down in an entire neighbourhood of each β simultaneously — closing this gap.

Connection to the LAD slides:
In *both* the generic and degenerate designs [Slides p. 13]–[Slides p. 15], the data are i.i.d., so the LLN applies and step (a) holds in *both* designs. This is why the algorithmic discrepancy eventually shrinks even in the degenerate case [Slides p. 21]: μ_n does converge to μ. The failure in the degenerate design is entirely in step (b), not step (a).

4 Step (b): Asymptotic Identification

▷ **Handwritten Notes** (what the professor said)

We require μ to satisfy a strong condition: not only must it be uniquely minimised at β_0, but β_0 must be **"well-separated"** from the other β's.
(Note: This is stronger than the simple identification condition we previously requested from μ_n^*.)
The plot below illustrates this concept. The global minimum at β_0 is strictly lower and clearly separated from any other local minima or flat asymptotes that the curve might have as it extends through the parameter space.

◇ **Student's Notes**

Formal definition of well-separation:
β_0 is **well-separated** in μ iff for every $\varepsilon > 0$ there exists $\delta > 0$ such that:

$$\inf_{\|\beta-\beta_0\|>\varepsilon} \mu(\beta) \;\geq\; \mu(\beta_0) + \delta.$$

Why "unique minimum" alone is not enough:
A function can have a unique global minimum yet fail well-separation if its tail asymptotes toward the minimum value. In that case $\delta \to 0$ as $\varepsilon \to \infty$ — no uniform gap exists — and the minimiser of μ_n can drift arbitrarily far from β_0.

The two contrasting cases:

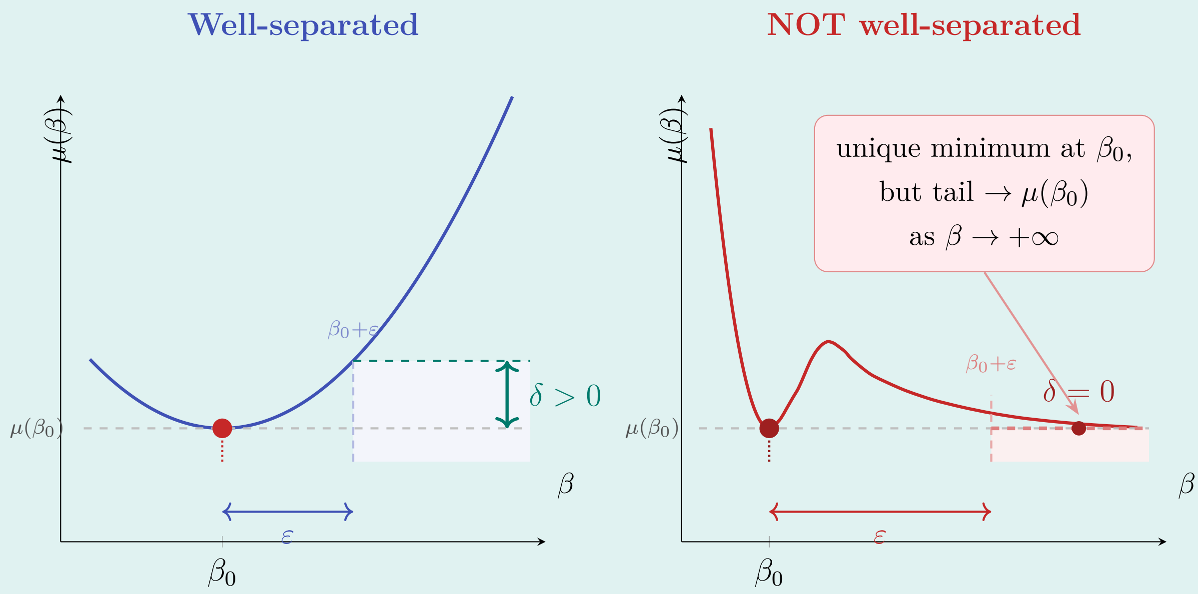

Figure 2: ***Left (indigo) — well-separated:*** *the parabola rises uniformly; for any $\varepsilon > 0$ the infimum of μ outside the ε-ball (teal dashed) lies strictly above $\mu(\beta_0)$, giving $\delta > 0$.* ***Right (rose) — not well-separated:*** *the minimum at β_0 is unique, but the tail decays toward $\mu(\beta_0)$; the infimum outside the ε-ball* collapses onto $\mu(\beta_0)$ *(rose dashed overlay), so no positive δ exists.*

5 The Professor's Illustrative Diagram

▷ **Handwritten Notes** (what the professor said)

The plot below illustrates the well-separated minimum concept. The global minimum at β_0 is strictly lower and clearly separated from any other local minima or flat asymptotes that the curve might have as it extends through the parameter space $\Theta = [0, +\infty)$.

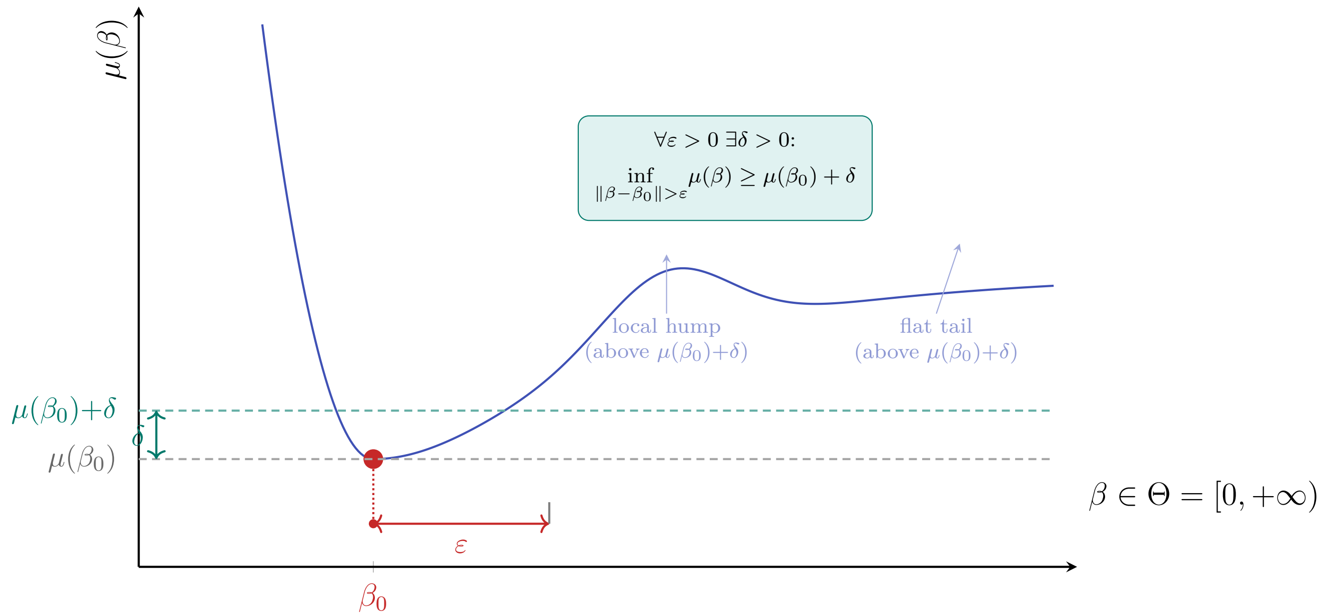

Figure 3: The professor's illustrative diagram. $\mu(\beta)$ has a unique global minimum at β_0 (rose dot). The gray dashed floor marks $\mu(\beta_0)$; the teal dashed line marks $\mu(\beta_0) + \delta$. Both the local hump ($\beta \approx 4.5$) and the flat tail ($\beta \to \infty$) lie above the teal line, so β_0 is well-separated.

6 Connection to the LAD Slides

▷ **Handwritten Notes** (what the professor said)

This theoretical requirement of a "well-separated minimum" directly explains the algorithmic behaviour observed in the LAD Monte Carlo experiments.
Generic design: The population objective has a unique, well-separated minimum at β_0. The asymptotic identification condition holds, so the sample estimator converges cleanly. Both algorithms (`highs-ds` and `highs-ipm`) agree perfectly.
Degenerate (multi-valued) design: Because the median of $\varepsilon_{(i)}$ is not unique, the population objective $\mu(\beta)$ develops "flat regions" at its minimum. The global minimum is **not well-separated**. The asymptotic identification condition completely fails. This is exactly why:

- The problem loses population identification.
- Simplex and interior-point return different results.
- The estimation error stays high even as $n \to +\infty$.

▷ Application to the LAD Slides

The theorem explains every slide result:

Lect. 10 concept	Generic design [Slides p. 13]	Degenerate design [Slides p. 14]
(a) Locally uniform convergence	✓ LLN holds (Laplace errors)	✓ LLN holds (discrete errors)
(b) Well-separated min	✓ unique median, $f(0) > 0$	× **fails**: flat region
Estimation error [Slides p. 20]–[Slides p. 21] ($c/\sqrt{n}$ decay) (stays large)	0.277 → 0.038 ≈ 0.42–0.45	
Algorithmic discrepancy [Slides p. 22], [Slides p. 24] (unique LP vertex) (shrinks but $\neq 0$)	$= 0$ for all n 0.180 → 0.014	
Consistency	✓ $\hat{\beta}_n \xrightarrow{p} \beta_0$	× Fails

Why discrepancy shrinks but error stays large [Slides p. 21]: Condition (a) holds in both designs, so μ_n concentrates around μ. As n grows the flat region of the *sample* criterion narrows ⇒ solvers pick increasingly similar points. But the *population* flat region is permanent ⇒ estimation error never vanishes.

Generic: (a) ✓ + (b) ✓ ⇒ consistent, solvers agree.
Degenerate: (a) ✓ + (b) × ⇒ inconsistent, solvers disagree.

◇ Student's Notes

Side-by-side: what $\mu(\beta)$ looks like in each design

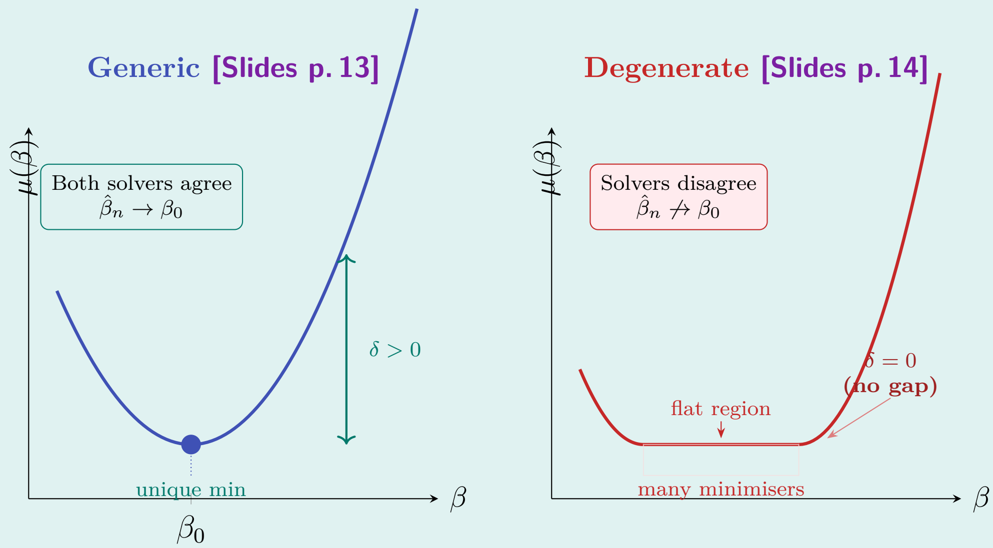

Figure 4: ***Left (generic):*** *smooth bowl, $\delta > 0$, unique LP vertex, both solvers agree [Slides p. 20], [Slides p. 22].* ***Right (degenerate):*** *flat bottom (shaded), $\delta = 0$, LP optimal set is an entire face, solvers return different points [Slides p. 21], [Slides p. 24].*

How the lectures build on each other:

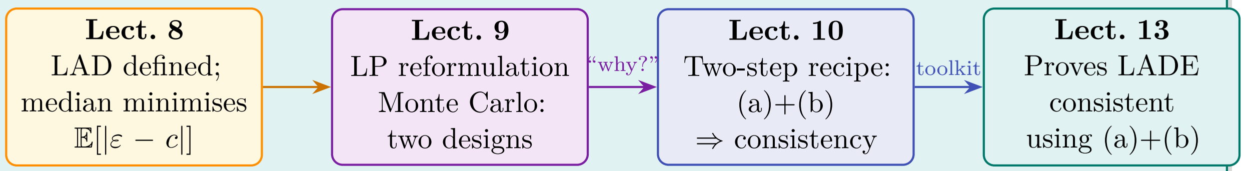

Figure 5: The lecture arc. Lect. 9 shows the phenomenon empirically; Lect. 10 provides the theorem; Lect. 13 applies the theorem to prove LADE consistency.

Quick-Reference Summary

Key Result

The Two-Step Consistency Recipe for M-Estimators

$\hat{\beta}_n \in \arg\min_{\beta\in\Theta} \mu_n(\beta) \xrightarrow{p} \beta_0$ follows from:

(a) Locally uniform convergence: $\sup_{\beta^*\in O_\beta} |\mu_n(\beta^*) - \mu(\beta^*)| \xrightarrow{p} 0$ for all β.

(b) Asymptotic identification (well-separated min): $\forall\varepsilon > 0\ \exists\delta > 0 : \inf_{\|\beta-\beta_0\|>\varepsilon} \mu(\beta) \geq \mu(\beta_0) + \delta$.

⋄ Student's Notes

Summary table:

Topic	What was established
Main obstacle	No closed form for $\hat{\beta}_n$; study μ_n instead
Step (a)	Locally uniform: $\sup_{O_\beta} \|\mu_n - \mu\| \xrightarrow{p} 0$; delivered by LLN
Step (b)	Well-separated min: $\inf_{\|\beta-\beta_0\|>\varepsilon} \mu \geq \mu(\beta_0) + \delta > 0$
LAD generic [Slides p. 13]	Both (a)+(b) hold ⇒ consistent, solvers agree [Slides p. 20]
LAD degenerate [Slides p. 14]	(a) holds, (b) fails ⇒ inconsistent, solvers disagree [Slides p. 21]
Master role	Lectures 11, 12, 13 verify (a)+(b) for OLS, IV, LAD

Cross-references to the LAD slides:

Slide	Content	Lect. 10 connection
[Slides p. 3]–[Slides p. 5]	LP reformulation of LAD	No closed form ⇒ need two-step recipe
[Slides p. 8]–[Slides p. 11]	Simplex vs. IPM algorithms	Selection rule within arg min set
[Slides p. 13]	Generic design	Conditions for (a)+(b) to hold
[Slides p. 14]–[Slides p. 15]	Degenerate design	Why (b) fails: non-unique median
[Slides p. 20]	Generic MC results	Empirical confirmation: consistent
[Slides p. 21]	Degenerate MC results	Empirical confirmation: (b) fails
[Slides p. 25]–[Slides p. 26]	Objective+geometry+algorithm	Three dimensions of Lecture 2