Lecture 13: Consistency of the LADE & Asymptotic Normality Basics

Econometrics 2 — *From LAD Consistency to the Taylor Expansion Framework*

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

Semester: Spring 2026

▷ Amber = Handwritten Notes (professor's words) ◇ Teal = Student's Notes

Recall from Earlier Lectures

Where we are in the course:

- **Lecture 8:** LAD estimator defined; proved $\mathrm{Med}(\varepsilon) = \arg\min_c \mathbb{E}[|\varepsilon - c|]$; showed $M_n^*(\beta)$ is uniquely minimised at β_0 under $\mathrm{Med}(\varepsilon \mid X) = 0$.
- **Lectures 10–11:** Two-step consistency recipe: (a) $M_n \xrightarrow{p} M$ locally uniformly, (b) M well-separated at β_0. Applied to OLS.
- **Lecture 12:** Same recipe applied to IV/GMM.

This lecture applies the recipe to LAD, proves convexity of the limit function (replacing the Cholesky/rank argument used for OLS/IV), then introduces the *general three-assumption framework* for asymptotic normality of M-estimators via Taylor expansion.

1 Assumptions for LADE Consistency

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

We investigate the consistency of the Least Absolute Deviations Estimator (LADE). We impose:

A1. The pairs $(\varepsilon_{(i)}, X_{(i)})$ are i.i.d. in i.

A2. $\mathbb{E}(|\varepsilon_{(1)}|)$ and $\mathbb{E}(\|X_{(1)}\|)$ exist.

A3. $\mathrm{Med}(\varepsilon_{(1)} \mid X_n) = 0$.

A4. The conditional distribution of $\varepsilon_{(1)} \mid X_n$ has a density f with $f(0) > 0$.

A5. $\mathrm{rank}(X_n) = p$.

Under A1–A5, the sample objective $M_n(\beta) = \frac{1}{n}\sum_{i=1}^n |Y_{(i)} - X_{(i)}\beta|$ converges locally uniformly in probability to:

$$M(\beta) = \mathbb{E}\big(M_n^*(\beta)\big), \qquad M_n^*(\beta) = \mathbb{E}\Big(\big|\varepsilon_{(i)} + X_{(i)}(\beta_0 - \beta)\big| \;\Big|\; X_n\Big).$$

◇ Student's Notes

How does LAD differ from OLS/IV at each step?

Aspect	OLS (L. 11)	IV (L. 12)	LAD (L. 13)
Objective M_n	$\frac{1}{n}\|Y-X\beta\|^2$	$\frac{1}{n^2}(Y-X\beta)'ZWZ'(\cdots)$	$\frac{1}{n}\sum_i \lvert Y_i - X_i'\beta\rvert$
Population condition	$\mathbb{E}[\varepsilon \mid X] = 0$	$\mathbb{E}[\varepsilon \mid Z] = 0$	$\text{Med}(\varepsilon \mid X) = 0$
Limit shape	Quadratic (smooth)	Quadratic (smooth)	Convex, piecewise linear
Unique-min proof	p.d. of $Q_{X'X}$	Cholesky: $Q'_{Z'X}WQ_{Z'X} \succ 0$	Convexity + unique median
Differentiable?	Yes (C^∞)	Yes (C^∞)	No (kinks at $Y_i = X_i'\beta$)

Role of each assumption:

Assumption	What it does in the proof
A1 (i.i.d.)	Enables Kolmogorov's LLN for $M_n(\beta) \xrightarrow{p} M(\beta)$, first pointwise then locally uniformly
A2 (moments)	Guarantees $M(\beta) = \mathbb{E}[\lvert\varepsilon + X(\beta_0 - \beta)\rvert] < \infty$ for every β
A3 (median zero)	From Lecture 8: $c = 0$ uniquely minimises $\mathbb{E}[\lvert\varepsilon - c\rvert \mid X]$, so β_0 minimises each $M_n^*(\beta)$
A4 ($f(0) > 0$)	Ensures the conditional median is *unique* (Lecture 8 uniqueness criterion); without this, M can have flat valleys
A5 (full rank)	$X(\beta_0 - \beta) = 0 \;\Rightarrow\; \beta = \beta_0$: no other parameter can mimic the true one

2 The Unique Minimum of the Limit Function

▷ Handwritten Notes (what the professor said)

We use $\min \mathbb{E}(\cdot) \geq \mathbb{E}(\min \cdot)$. Applying to M_n^*:

$$\min_{\beta\in\Theta} M(\beta) = \min_{\beta\in\Theta} \mathbb{E}(M_n^*(\beta)) \;\geq\; \mathbb{E}\left(\min_{\beta\in\Theta} M_n^*(\beta)\right) = \mathbb{E}(M_n^*(\beta_0)) = M(\beta_0).$$

Since $\beta_0 \in \Theta$, also $\min_{\beta\in\Theta} M(\beta) \leq M(\beta_0)$, giving equality.
Uniqueness: Suppose $\beta^* \neq \beta_0$ also minimises M. From Lecture 8, $M_n^*(\beta^*) > M_n^*(\beta_0)$ pointwise:

$$M(\beta^*) = \mathbb{E}(M_n^*(\beta^*)) > \mathbb{E}(M_n^*(\beta_0)) = M(\beta_0) = \min_{\beta\in\Theta} M(\beta).$$

Contradiction. Hence β_0 is the **unique minimiser**.

◇ Student's Notes

Visualising M, M_n^*, and M_n:

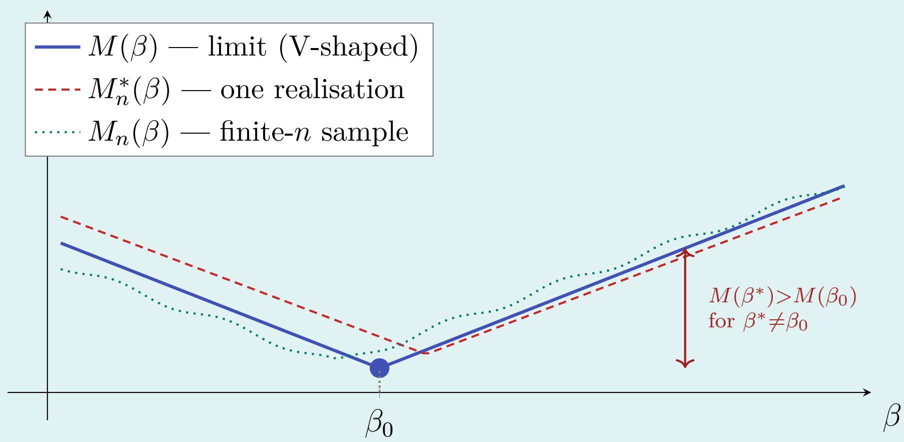

Figure 1: The limit $M(\beta)$ (indigo) is V-shaped with unique minimum at β_0. Individual realisations $M_n^*(\beta)$ (rose dashed) and the sample $M_n(\beta)$ (teal dotted) converge to M as $n\to\infty$. The rose arrow shows the gap $M(\beta^*)>M(\beta_0)$ that drives the contradiction argument.

Logical chain:

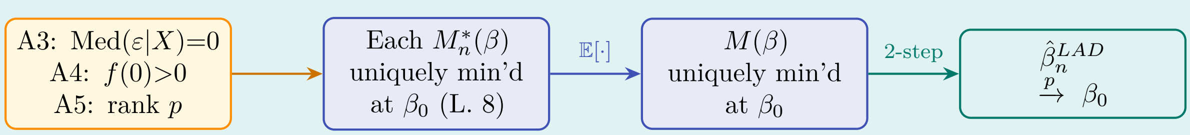

Figure 2: Logical chain for LADE consistency.

3 Convexity of the Limit Function

▷ Handwritten Notes (what the professor said)

$M(\beta)$ is convex iff for all $\lambda \in [0,1]$ and all $\beta_1, \beta_2 \in \mathbb{R}^p$: $M(\lambda\beta_1 + (1-\lambda)\beta_2) \leq \lambda M(\beta_1) + (1-\lambda)M(\beta_2)$.

Proof:

$$\begin{aligned}
&M\big(\lambda\beta_1 + (1-\lambda)\beta_2\big) \\
&\quad = \mathbb{E}\Big(\big|\lambda\,\underbrace{(\varepsilon_{(1)} + X_{(1)}(\beta_0 - \beta_1))}_{A} + (1-\lambda)\,\underbrace{(\varepsilon_{(1)} + X_{(1)}(\beta_0 - \beta_2))}_{B}\big|\Big) \\
&\quad \overset{|A+B| \leq |A|+|B|}{\leq} \lambda\,\mathbb{E}(|A|) + (1-\lambda)\,\mathbb{E}(|B|) \;=\; \lambda M(\beta_1) + (1-\lambda)M(\beta_2).
\end{aligned}$$

Since M is convex and uniquely minimised at β_0, it is **strictly convex**, β_0 is well-separated, and $\hat{\beta}_n^{LAD} \xrightarrow{p} \beta_0$.

■ Proof / Derivation

Why "convex + unique minimum" ⇒ strictly convex

A convex function can satisfy the chord inequality with *equality* only if it is affine on the segment $[\beta_1, \beta_2]$. But an affine segment means every point on it achieves the same objective value — contradicting the fact that β_0 is the *unique* minimiser. So the chord inequality is strict for all $\beta_1 \neq \beta_2$.

Well-separation then follows automatically: for a strictly convex function with unique minimum at β_0 over convex Θ, $\inf_{\|\beta-\beta_0\|>\varepsilon} M(\beta) \geq M(\beta_0) + \delta$ for some $\delta > 0$, because convexity rules out flat tails. □

◇ Student's Notes

The chord lies strictly above the curve:

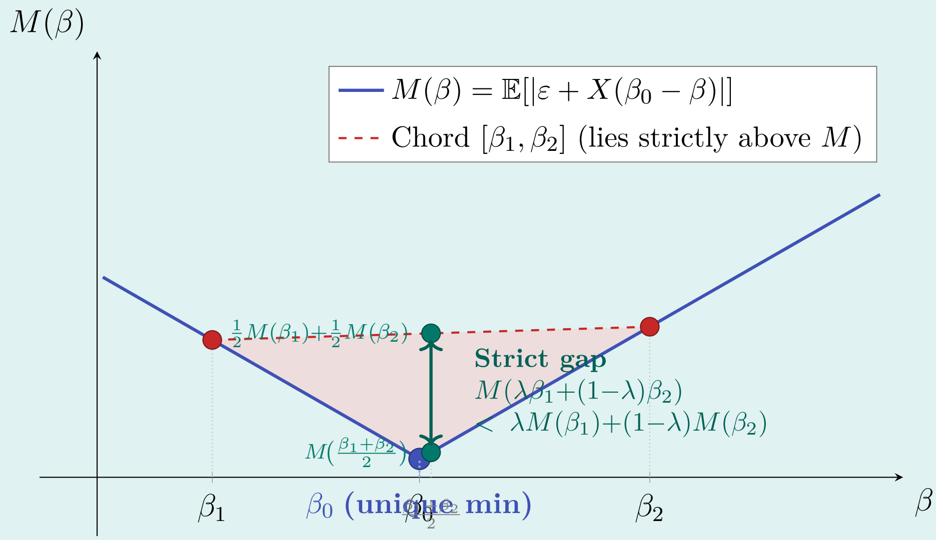

Figure 3: Convexity of $M(\beta)$. The chord (rose dashed) between any two points on the curve lies strictly above M (shaded region). The teal arrow marks the gap between the function value and the chord value at a midpoint—strict convexity in action.

Contrast with OLS/IV: For OLS and IV the limit is a *smooth quadratic* and strict convexity comes from positive definiteness of a curvature matrix. For LAD the limit is *piecewise linear* (V-shaped) and strict convexity comes from the uniqueness of the conditional median. Both yield the same conclusion: well-separated minimum at β_0.

Key Result

Weak Consistency of the LADE

Under A1–A5:

$$M(\beta) = \mathbb{E}\Big[\big|\varepsilon_{(1)} + X_{(1)}(\beta_0 - \beta)\big|\Big]$$

is **strictly convex** (triangle inequality + unique median) with a unique, well-separated minimum at β_0. Therefore:

$$\boxed{\hat{\beta}_n^{LAD} \xrightarrow{p} \beta_0}$$

4 Topology of Θ: Interior, Closure, Boundary

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

Having established weak consistency, we want sufficient conditions for rate $\sqrt{n}$ and asymptotic normality. This requires the geometry of $\Theta \subset \mathbb{R}^p$.

- **Interior:** $\mathrm{Int}(\Theta)$ — largest open subset of Θ.
- **Closure:** $\bar{\Theta}$ — smallest closed set containing Θ.
- **Boundary:** $\partial\Theta = \bar{\Theta} \setminus \mathrm{Int}(\Theta)$.

Example ($p = 1$): $\Theta = (a, b] \;\Rightarrow\; \mathrm{Int}(\Theta) = (a, b),\ \bar{\Theta} = [a, b],\ \partial\Theta = \{a, b\}$.

◇ **Student's Notes**

Why the geometry matters:
The key forthcoming assumption is $\beta_0 \in \mathrm{Int}(\Theta)$. If β_0 lies on $\partial\Theta$, the minimiser may be a *corner solution* where $\nabla\mu_n \neq 0$, breaking the Taylor expansion.

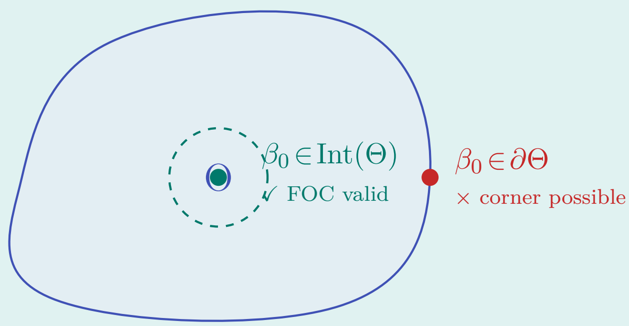

Figure 4: Interior condition. The teal dot (with dashed open ball) is in $\mathrm{Int}(\Theta)$*: the FOC holds and the Taylor expansion is valid. The rose dot on* $\partial\Theta$ *may be a corner solution where* $\nabla\mu_n \neq 0$.

5 General Framework for Asymptotic Normality

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

To prove asymptotic normality for a general M-estimator we use three assumptions:

Assumption 1 (Interior): $\beta_0 \in \text{Int}(\Theta)$.
Since $\hat{\beta}_n \xrightarrow{p} \beta_0$, this implies $\hat{\beta}_n \in \text{Int}(\Theta)$ w.p. $\to 1$.

Assumption 2 (Smoothness): $\mu_n(\beta)$ is twice continuously differentiable (C^2).
Together, $1 + 2$ yield the FOC w.p. $\to 1$:

$$\frac{\partial \mu_n(\hat{\beta}_n)}{\partial \beta} = 0_{p\times 1}.$$

Taylor-expanding around β_0:

$$0 = \frac{\partial \mu_n(\beta_0)}{\partial \beta} + \frac{\partial^2 \mu_n(\beta_n^*)}{\partial \beta\, \partial \beta'}(\hat{\beta}_n - \beta_0),$$

where β_n^* lies between $\hat{\beta}_n$ and β_0, so $\beta_n^* \xrightarrow{p} \beta_0$.

Assumption 3 (Score CLT):

$$\sqrt{n}\, \frac{\partial \mu_n(\beta_0)}{\partial \beta} \xrightarrow{d} N(0_{p\times 1}, V), \qquad V \succ 0.$$

◇ **Student's Notes**

From the FOC to the sandwich formula:
Define $H_n := \frac{\partial^2 \mu_n(\beta_n^*)}{\partial \beta\, \partial \beta'}$. By LLN and $\beta_n^* \xrightarrow{p} \beta_0$: $H_n \xrightarrow{p} H$ (the population Hessian). Rearranging:

$$\sqrt{n}\,(\hat{\beta}_n - \beta_0) = -H_n^{-1} \cdot \underbrace{\sqrt{n}\, \frac{\partial \mu_n(\beta_0)}{\partial \beta}}_{\xrightarrow{d} N(0,V)}.$$

Slutsky's theorem ($H_n^{-1} \xrightarrow{p} H^{-1}$) gives:

$$\sqrt{n}\,(\hat{\beta}_n - \beta_0) \xrightarrow{d} N\big(0,\, \underbrace{H^{-1} V H^{-1}}_{\text{sandwich}}\big).$$

Convergence summary:

Object	Limit	Tool
$H_n = \nabla^2 \mu_n(\beta_n^*)$	H (population Hessian)	LLN + CMT
$\sqrt{n}\,\nabla \mu_n(\beta_0)$	$N(0, V)$	CLT (Assumption 3)
$\sqrt{n}(\hat{\beta}_n - \beta_0)$	$N(0, H^{-1} V H^{-1})$	Slutsky

Flow diagram:

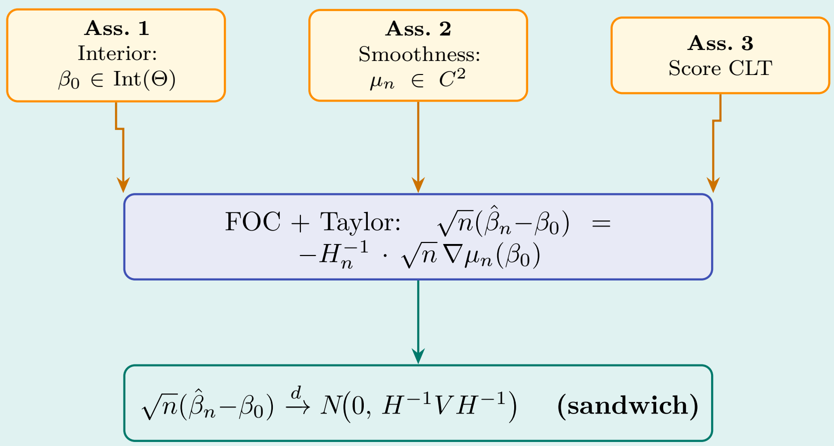

Figure 5: The three-assumption framework. Assumptions 1–2 enable the FOC and Taylor expansion. Assumption 3 provides the score CLT. Slutsky combines $H_n^{-1} \xrightarrow{p} H^{-1}$ with the CLT to give the sandwich distribution.

! Watch Out

Assumption 2 (smoothness) fails for LAD.
The LAD objective $M_n(\beta) = \frac{1}{n}\sum_i |Y_i - X_i'\beta|$ is *not differentiable* at any β where some residual $Y_i - X_i'\beta = 0$. The Taylor expansion cannot be applied directly.

Two routes to normality without smoothness:

1. **Knight's identity:** An algebraic identity $|a - b| - |a| = -b\,\mathrm{sign}(a) + 2\int_0^b \big(\mathbf{1}_{a \le s} - \mathbf{1}_{a \le 0}\big) ds$ that replaces derivatives with indicator functions.

2. **Convexity lemma (Pollard):** For convex objectives, pointwise convergence implies convergence of minimisers, bypassing differentiability entirely.

Both are developed in **Lecture 14**.

Quick-Reference Summary

◇ Student's Notes

Lecture 13 at a glance:

Topic	What was accomplished
LADE assumptions	A1–A5 listed; role of each explained
Unique minimum of M	Contradiction via $M_n^*(\beta^*) > M_n^*(\beta_0)$ from Lecture 8
Convexity of M	Triangle ineq. + $\mathbb{E}$ linearity; strict convexity from unique min
LADE consistency	$\hat{\beta}_n^{LAD} \xrightarrow{p} \beta_0$
Topology of Θ	Interior/closure/boundary; interior condition for normality
Normality framework	3 assumptions: interior, smoothness, score CLT; $\Rightarrow$ sandwich $N(0, H^{-1}VH^{-1})$
LAD bottleneck	Smoothness fails; Knight/Pollard in Lecture 14

Key formulas:

Object	Expression
LAD objective	$M_n(\beta) = \frac{1}{n}\sum_{i=1}^n \lvert Y_i - X_i'\beta\rvert$
Limit function	$M(\beta) = \mathbb{E}[\lvert\varepsilon_{(1)} + X_{(1)}(\beta_0 - \beta)\rvert]$
Taylor expansion	$0 = \nabla\mu_n(\beta_0) + \nabla^2\mu_n(\beta_n^*)(\hat{\beta}_n - \beta_0)$
Sandwich	$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, H^{-1}VH^{-1})$
Score CLT	$\sqrt{n}\,\nabla\mu_n(\beta_0) \xrightarrow{d} N(0, V)$

Lectures 2–13 narrative arc:

Lectures	Theme
2–3	Model setup; population identification
4–5	Analogy principle; extremum estimators; asymptotic properties
6–7	Endogeneity; IV / GMM; just-identified case
8–9	LAD estimator; LP reformulation
10–12	Two-step consistency recipe; OLS and IV consistency
13	**LADE consistency (convexity); topology of Θ; asymptotic normality framework**