Econometrics 2

Lecture 13: Consistency of the LADE

& Asymptotic Normality Basics

AUEB | Spring Semester 2026

Taught by: Prof. S. Arvanitis

Notes transcribed & digitized by: T. Kourtalis

1. Conditions for the Consistency of the LAD Estimator

We are trying to investigate the issue of consistency for the Least Absolute Deviations Estimator (LADE). We establish the following assumptions:

1. The pairs $(\epsilon_{(i)}, X_{(i)})$ are independent and identically distributed (iid) with respect to i.
2. The expectations $\mathbb{E}(|\epsilon_{(1)}|)$ and $\mathbb{E}(||X_{(1)}||)$ exist.
3. $Med(\epsilon_{(1)}|X_n) = 0$.
4. The conditional distribution of $\epsilon_{(1)}|X_n$ has a density function, let us call it f, and $f(0) > 0$.
5. $rank(X_n) = p$.

Under these preconditions, we saw that the sample objective function:

$$M_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} |Y_{(i)} - X_{(i)}\beta|$$

converges locally uniformly in probability to a limit function $M(\beta)$.

By the Law of Iterated Expectations (L.I.E), this limit is:

$$M(\beta) = \mathbb{E}(M_n^*(\beta))$$

where $M_n^*(\beta) = \mathbb{E}\Big(|\epsilon_{(i)} + X_{(i)}(\beta_0 - \beta)|\Big|X_n\Big)$.

2. The Unique Minimum

We noted that because the general inequality $\min \mathbb{E}(\dots) \geq \mathbb{E}(\min \dots)$ holds, we will have:

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

And since $\beta_0 \in \Theta$, it belongs to the feasible set, therefore:

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

Question: Is β_0 the *unique* element of $\arg\min_{\beta \in \Theta} M(\beta)$?

Suppose it is not. Then there exists some $\beta^* \in \Theta$ with $\beta^* \neq \beta_0$ that also minimizes M. This would mean:

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

However, from our previous derivations regarding medians, we know that $M_n^*(\beta^*) > M_n^*(\beta_0)$. Therefore:

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

This is a contradiction. Thus, β_0 is the unique minimizer.

3. Convexity of the Objective Function

The function M as a function of $\beta \in \mathbb{R}^p$ will be convex if $\forall \lambda \in [0, 1]$ and for any $\beta_1, \beta_2 \in \mathbb{R}^p$:

$$M(\lambda\beta_1 + (1-\lambda)\beta_2) \leq \lambda M(\beta_1) + (1-\lambda)M(\beta_2)$$

Let us prove this by plugging the convex combination into our function:

$$\begin{aligned}
M(\lambda\beta_1 + (1-\lambda)\beta_2) &= \mathbb{E}\Big(|\epsilon_{(1)} + X_{(1)}\big(\beta_0 - (\lambda\beta_1 + (1-\lambda)\beta_2)\big)|\Big) \\
&= \mathbb{E}\Big(|\lambda\epsilon_{(1)} + (1-\lambda)\epsilon_{(1)} + X_{(1)}\big(\lambda\beta_0 + (1-\lambda)\beta_0 - \lambda\beta_1 - (1-\lambda)\beta_2\big)|\Big) \\
&= \mathbb{E}\Big(|\lambda\big(\epsilon_{(1)} + X_{(1)}(\beta_0 - \beta_1)\big) + (1-\lambda)\big(\epsilon_{(1)} + X_{(1)}(\beta_0 - \beta_2)\big)|\Big)
\end{aligned}$$

Using the triangle inequality ($|A + B| \leq |A| + |B|$):

$$\begin{aligned}
&\leq \mathbb{E}\Big(\lambda|\epsilon_{(1)} + X_{(1)}(\beta_0 - \beta_1)| + (1-\lambda)|\epsilon_{(1)} + X_{(1)}(\beta_0 - \beta_2)|\Big) \\
&= \lambda\mathbb{E}\Big(|\epsilon_{(1)} + X_{(1)}(\beta_0 - \beta_1)|\Big) + (1-\lambda)\mathbb{E}\Big(|\epsilon_{(1)} + X_{(1)}(\beta_0 - \beta_2)|\Big) \\
&= \lambda M(\beta_1) + (1-\lambda)M(\beta_2)
\end{aligned}$$

Therefore, M is a convex function of β. Because it is uniquely minimized at β_0, it must be **strictly convex**. Consequently, β_0 is the unique minimizing point of M and is well-separated. Because of our general asymptotic theory, this implies:

$$\beta_n \xrightarrow{P} \beta_0$$

4. Topological Concepts (Parameter Space)

Continuing, given the weak consistency, we want to establish sufficient conditions that will ensure the estimator has the classical convergence rate $\sqrt{n}$ and is asymptotically normal. We will use higher-order conditions, which will require understanding the geometry of $\Theta \subset \mathbb{R}^p$.

- **Interior:** The interior of Θ (denoted $Int(\Theta)$) is the largest open subset of Θ.
- **Closure:** The closure of Θ (denoted $\bar{\Theta}$) is the smallest closed subset of $\mathbb{R}^p$ that contains Θ.
- **Boundary:** The difference $\bar{\Theta} - Int(\Theta)$ is the boundary of Θ.

Example ($p = 1$): If $\Theta = (a, \beta]$, then $Int(a, \beta] = (a, \beta)$, the closure is $\bar{\Theta} = [a, \beta]$, and the boundary is the set $\{a, \beta\}$.

5. General Conditions for Asymptotic Normality

To prove asymptotic normality for a general M-estimator, we introduce the following assumptions:

1. Interior Condition: $\beta_0 \in Int(\Theta)$

Because our estimator is weakly consistent ($\beta_n \xrightarrow{P} \beta_0$), Assumption 1 implies that β_n will also fall inside the interior of Θ with a probability that converges to 1.

2. Smoothness: The objective function $\mu_n(\beta)$ is twice continuously differentiable. Assumptions 1 and 2 imply that, with probability approaching 1, β_n satisfies the First-Order Conditions (since it is an interior minimum of a smooth function). Namely:

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

Because of Assumption 2, we can use a Taylor expansion on the left side of this system around the true parameter β_0:

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

Where β_n^* lies on the line segment connecting β_n and β_0, meaning $||\beta_n^* - \beta_0|| < ||\beta_n - \beta_0||$. Since $||\beta_n - \beta_0|| \xrightarrow{P} 0$ (due to consistency), we have $||\beta_n^* - \beta_0|| \xrightarrow{P} 0$. Therefore, $\beta_n^* \xrightarrow{P} \beta_0$.

3. Asymptotic Normality of the Score:

$$n^{1/2}\frac{\partial \mu_n(\beta_0)}{\partial \beta} \xrightarrow{d} Z \sim N(0_{p\times 1}, V)$$

Where V is a positive definite $p \times p$ matrix. *(Generally, this matrix will be unknown and we may need to estimate it).*