Econometrics 2

Lecture 8: The LAD Estimator

Medians, Absolute Deviations, and the Linear Model

AUEB | Spring Semester 2026
Taught by: Prof. S. Arvanitis
Notes transcribed & digitized by: T. Kourtalis

April 25, 2026

1. Linear Model and the LAD Estimator

The Least Absolute Deviations Estimator (LADE) results from the minimization of:

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

Question: What modification of the linear model could support such an objective function?

The modification will be related to the **median** of the distribution of $\epsilon_{(i)}$, where:

$$\epsilon_n = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

2. Defining the Median

Let the random variable $\epsilon \sim \mathbb{P}$ (where $\mathbb{P}$ is a probability distribution on $\mathbb{R}$). The cumulative distribution function (CDF) of $\mathbb{P}$ is $F : \mathbb{R} \to \mathbb{R}$, defined as:

$$F(x) := \mathbb{P}(\epsilon \leq x)$$

The median of $\mathbb{P}$ (or equivalently of ϵ) is any real number ζ such that:

$$F(\zeta) = \mathbb{P}(\epsilon \leq \zeta) = \frac{1}{2} \iff \mathbb{P}(\epsilon > \zeta) = \frac{1}{2}$$

Example (Discrete Distribution)

Suppose ϵ takes values:

$$\epsilon = \begin{cases} -1, & \text{with probability } 1/2 \\ 1, & \text{with probability } 1/2 \end{cases}$$

The CDF is:

$$F(x) = \begin{cases} 0, & x < -1 \\ 1/2, & -1 \leq x < 1 \\ 1, & 1 \leq x \end{cases}$$

In this specific case, *any* value $\zeta \in [-1, 1]$ is a median of this distribution.

Sufficient Condition for Uniqueness

A sufficient condition for the median $Med(\epsilon)$ to be uniquely defined is for $\mathbb{P}$ to have a probability density function, say f, and that $f(Med(\epsilon)) > 0$.

3. Minimizing Expected Absolute Deviation

Let us assume that the expectation $\mathbb{E}(\epsilon)$ exists (which holds $\iff \mathbb{E}(|\epsilon|) < +\infty$).

This implies that the function:

$$\psi(c) := \mathbb{E}(|\epsilon - c|)$$

is well-defined (i.e., $\psi(c) \in \mathbb{R} \quad \forall c \in \mathbb{R}$). Therefore, we can ask whether there is some constant c at which ψ is minimized.

We know that:

$$|\epsilon - c| = \begin{cases} \epsilon - c, & \epsilon \geq c \\ c - \epsilon, & \epsilon < c \end{cases}$$

We will ignore the specific point $c = \epsilon$ for which we do not have differentiability. The derivative with respect to c is:

$$\frac{\partial |\epsilon - c|}{\partial c} = \begin{cases} -1, & \epsilon > c \\ 1, & \epsilon < c \end{cases}$$

4. First-Order Conditions (FOC)

Assuming $\mathbb{P}$ has a continuous probability function, we can ignore the single point of non-differentiability. By the Dominated Convergence Theorem, we can pass the derivative inside the expectation:

$$\frac{\partial \mathbb{E}(|\epsilon - c|)}{\partial c} = \mathbb{E}\left(\frac{\partial |\epsilon - c|}{\partial c}\right) = \mathbb{E} \begin{cases} -1, & \epsilon > c \\ 1, & \epsilon < c \end{cases}$$

$$= -1 \cdot \mathbb{P}(\epsilon > c) + 1 \cdot \mathbb{P}(\epsilon < c)$$

Therefore, the First-Order Condition to find the critical point c^* that minimizes the function is:

$$\left.\frac{\partial \mathbb{E}(|\epsilon - c|)}{\partial c}\right|_{c=c^*} = 0$$

$$\iff -\mathbb{P}(\epsilon > c^*) + \mathbb{P}(\epsilon < c^*) = 0$$

$$\iff \mathbb{P}(\epsilon < c^*) = \mathbb{P}(\epsilon > c^*)$$

Because we assume $\mathbb{P}$ has a continuous density function (meaning $\mathbb{P}(\epsilon = c^*) = 0$), we can write:

$$\mathbb{P}(\epsilon \leq c^*) = \mathbb{P}(\epsilon \geq c^*)$$

Using the complement rule $\mathbb{P}(\epsilon \geq c^*) = 1 - \mathbb{P}(\epsilon \leq c^*)$, we get:

$$\mathbb{P}(\epsilon \leq c^*) = 1 - \mathbb{P}(\epsilon \leq c^*)$$

$$\iff 2\mathbb{P}(\epsilon \leq c^*) = 1$$

$$\iff \mathbb{P}(\epsilon \leq c^*) = \frac{1}{2}$$

This proves that the constant c^* that minimizes the expected absolute deviation is exactly the **median** of the distribution.

5. Returning to the Linear Model

Let us return to the linear model where we have our correct specification:

$$Y_n = X_n\beta_0 + \epsilon_n \iff Y_{(i)} = X_{(i)}\beta_0 + \epsilon_{(i)} \quad \forall i = 1, \ldots, n$$

We will make the following assumptions:

1. We assume the distribution of $\epsilon_{(i)}$ given X_n is independent of i (Homogeneity).
2. We assume that the conditional median is zero:

$$Med(\epsilon_{(i)}|X_n) = 0$$

3. We assume that the conditional distribution of $\epsilon_{(i)}|X_n$ has a density function, say f^*, such that $f^*(0) > 0$.

Based on what we proved previously, these assumptions imply that:

$$Med(Y_{(i)}|X_n) = X_{(i)}\beta_0 \quad \forall i = 1, \ldots, n$$

Consequently, searching for β_0 is equivalent to searching for the conditional median $Med(Y_{(i)}|X_n)$ for all $i = 1, \ldots, n$.

6. Analyzing the Objective Function

Let us examine the expected value of our proposed absolute deviation objective function:

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}|Y_{(i)} - X_{(i)}\beta|\middle|X_n\right)$$

Using the linearity of expectations, we can bring the sum outside:

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(|Y_{(i)} - X_{(i)}\beta|\middle|X_n\right)$$

Now, we substitute the true data-generating process $Y_{(i)} = X_{(i)}\beta_0 + \epsilon_{(i)}$:

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(|X_{(i)}\beta_0 + \epsilon_{(i)} - X_{(i)}\beta|\middle|X_n\right)$$

By factoring out $X_{(i)}$, we get:

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(|\epsilon_{(i)} - X_{(i)}(\beta - \beta_0)|\middle|X_n\right) \quad (*)$$

7. Minimization and the True Parameter

Let us define a function ψ inside the expectation:

$$\psi(c) = \mathbb{E}\left(|\epsilon_{(i)} - c|\middle|X_n\right)$$

Based on our assumptions (specifically that the conditional median of ϵ is zero), we know that $\psi(c)$ is uniquely minimized at $c = 0$.

Looking at equation $(*)$, the term acting as "c" is $X_{(i)}(\beta - \beta_0)$. Therefore, we know that each individual term in the sum is minimized at any β^* for which:

$$X_{(i)}(\beta^* - \beta_0) = 0$$

(Note: One obvious solution is $\beta^* = \beta_0$).

Furthermore, when the full rank condition holds ($rank(X_n) = p$), then the *entire sum* is minimized uniquely at β_0.

Let's define the theoretical expected objective function as $\mu_n^*(\beta)$:

$$\mu_n^*(\beta) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(|\epsilon_{(i)} - X_{(i)}(\beta - \beta_0)|\middle|X_n\right)$$

We have shown that:

$$\min_{\beta}\mu_n^*(\beta) = \frac{1}{n}\sum_{i=1}^{n}\min_{\beta}\mathbb{E}(\dots)$$

(Mathematical property note: generally $\min(A + B) \geq \min(A) + \min(B)$, but here the minimum is achieved simultaneously for all i at β_0).

8. The Sample LAD Estimator

The theoretical function $\mu_n^*(\beta)$ is not observable because it depends on the unknown true parameter β_0.

Therefore, it is approximated by its sample analog:

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

The M-estimator β_n defined by:

$$\beta_n \in \arg\min_{\beta \in \Theta} \mu_n(\beta)$$

is called the **Least Absolute Deviations (LAD) Estimator**.