Econometrics 2

Lecture 10: Asymptotic Theory of M-Estimators

(Lecture Comments & Notes)

AUEB | Spring Semester 2026

Taught by: Prof. S. Arvanitis

Notes transcribed & digitized by: T. Kourtalis

These notes cover the handwritten material from the lecture session. They serve as the theoretical foundation for the LAD behavior observed in the official slides of Prof. S. Arvanitis, available at:

Official LAD Slides (eClass)

---

1. Goals of the Asymptotic Theory of M-Estimators

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

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

Question: Are there sufficient conditions for the (weak) consistency of β_n?

2. The Main Obstacle and Methodology

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 behavior of the objective function itself as $n \to +\infty$.

The relevant methodology usually consists of the following steps:

(a) Achieving an "appropriate convergence" in probability of μ_n to a limit function (let's call it μ, which usually depends on β).

(b) If μ also satisfies some "asymptotic identification" condition, then this implies, among other things, that:

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

3. Regarding Step (a): Appropriate Convergence

One form of "appropriate convergence" in probability (though it is not the only one) is what is called **locally uniform convergence in probability** of μ_n to μ.

This means that for every possible value of β, we must be able to find a set of neighboring points, say O_β, such that the supremum of the absolute difference goes to zero in probability:

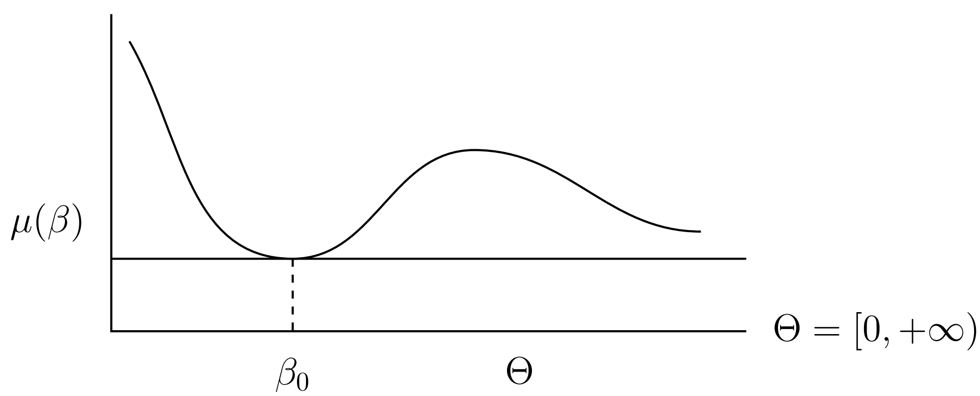
$$\sup_{\beta^* \in O_\beta} |\mu_n(\beta^*) - \mu(\beta^*)| \xrightarrow{P} 0$$

4. Regarding Step (b): Asymptotic Identification

Regarding the second conditions, we require μ to satisfy a strong condition: not only must it be uniquely minimized at β_0, but β_0 must also 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.



5. Connection to the LAD Slides (Application of the Theory)

This theoretical requirement of a "well-separated minimum" directly explains the algorithmic behavior observed in the LAD Monte Carlo experiments.

In the Generic Design: The population objective function has a unique, well-separated minimum at β_0. Because the asymptotic identification condition holds, the sample estimator converges cleanly to the true parameter. This is why both algorithms (`highs-ds` and `highs-ipm`) agree perfectly.

In the Degenerate (Multi-valued) Design: Because the median of the error term $\epsilon_{(i)}$ is not unique, the population objective function $\mu(\beta)$ develops "flat regions" at its minimum (arg min). Consequently, the global minimum is **not well-separated** (and not unique). The asymptotic identification condition completely fails. This is exactly why:

- The problem loses population identification.
- The simplex and interior-point algorithms return different results (algorithmic non-coincidence).
- The estimator fails to converge to a single true β_0, keeping the estimation error high even as $n \to +\infty$.