Econometrics 2

Lecture 5: General Properties of Estimators & Extremum Estimators

Asymptotic Behavior and M-Estimators

AUEB | Spring Semester 2026

Taught by: Prof. S. Arvanitis
Notes transcribed by: E. Hatija
Digitization & Comments by: T. Kourtalis

Reminder: General Properties of Estimators

We are concerned with recalling useful general properties of estimators within the framework of (semi-)parametric models.

Sample:

- $Z_n \to \Theta$

$\hookrightarrow$ provided it has a well-defined probability distribution.

In many cases, the exact distribution of the estimator (for a given n) is difficult (and in practice, impossible) to derive. We usually rely on "corresponding" asymptotic properties.

1. Weak Consistency

A primary property is weak consistency. An estimator β_n is weakly consistent iff:

$$\beta_n \xrightarrow{p} \beta_0$$

$\hookrightarrow$ Roughly speaking, this means that the probability that the distance between β_n and β_0 is strictly positive converges to 0.

$\to$ In most of the cases we will examine, consistency will arise from the appropriate convergence of the objective criterion to an asymptotic criterion, which, under some asymptotic identification condition, will be uniquely minimized at β_0. Useful tools in such cases will be the Laws of Large Numbers (LLN).

2. Rate of Convergence

A second asymptotic property we will examine, given weak consistency, is the rate of convergence. This represents a "sense of speed" at which β_n converges to β_0.

The rate of convergence will be a real sequence with term $\mathcal{F}(n)$ where $\mathcal{F}(n) \to +\infty$ as $n \to +\infty$, with the property that:

$$\mathcal{F}_n(\beta_n - \beta_0) \text{ asymptotically stabilizes to a well-defined random vector.}$$

Usually—and certainly in what we do—$\mathcal{F}(n) = \sqrt{n}$ due to the operation of some Central Limit Theorem (CLT).

3. Asymptotic Distribution

The third concept that will concern us is that of the asymptotic distribution. Given weak consistency and the rate of convergence $\mathcal{F}(n)$, we will have:

$$\mathcal{F}_n(\beta_n - \beta_0) \xrightarrow{d} Z \sim \text{some well-defined probability curve / limiting distribution}$$

In what we will see below, as we said previously:

$$Z \sim N(0_{p\times 1}, V)$$

$\hookrightarrow$ where V is a $p \times p$ variance-covariance matrix.

Therefore, we will have:

$$\sqrt{n}(\beta_n - \beta_0) \xrightarrow{d} Z \sim N(0_{p\times 1}, V)$$

and V will be called the asymptotic covariance matrix of β_n.

* To derive the rate of convergence, the asymptotic normality, and V, useful tools will be the Laws of Large Numbers, the Central Limit Theorem, and Taylor expansions.

Extremum Estimators (M-estimators)

- We work within a (semi-)parametric model with parameter space $\Theta \subseteq \mathbb{R}^p$. *(Goal: finding β_0).*
- Usually, we will assume that the model is well-specified (we will also examine some cases where this does not hold).
- We have at our disposal a function $M_n(\beta)$; this is observable and "approximates" (it depends on the sample) an unobservable function (let's say $M_n^*(\beta)$) which is uniquely minimized at β_0.

The extremum estimator → let's call it β_n (based on the above) will be defined as:

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

This yields a function $Z_n \to \Theta$, since:

$$Z \xrightarrow{\text{evaluated at some}} M_n(\beta) \implies \text{we then have } M_n(\beta) \text{ as a function of } \beta$$

$$\xrightarrow{\text{we minimize it w.r.t.}} \beta \in \Theta$$

Practical Difficulties in Minimization

$$\to \arg\min_{\beta\in\Theta} M_n(\beta) \in \Theta$$

In this case, it is practically possible to encounter various difficulties, e.g.:

(a) It is possible for "many values" of Z_n that the $\arg\min_{\beta\in\Theta} M_n(\beta)$ is empty. Usually, this won't happen if, for example, Θ is convex and M_n is a convex function of β for most values Z_n can take.

(b) It is possible that the $\arg\min_{\beta\in\Theta} M_n(\beta)$ has more than one element, so we would need to choose *(we won't deal directly with such cases)*.

(c) **Most Important!!** In many cases, the analytical optimization of M_n will be practically impossible. It must be done numerically. The choice of the respective algorithm can impact the properties of the estimator (e.g., linear programming, or Newton-Raphson type algorithms that use derivatives, etc.).

Example: The (Semi-)parametric Linear Model

As we have developed it:

$$M_n(\beta) = \frac{1}{n}(Y_n - X_n\beta)'(Y_n - X_n\beta)$$

Consequently, the extremum estimator resulting from the minimization over Θ of the above is the Ordinary Least Squares Estimator (OLSE):

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

- When $\Theta = \mathbb{R}^p$, the problem has an analytical solution. And we know that:

$$\beta_n = (X_n' X_n)^{-1} X_n' Y_n$$

(Note: requires $(X_n' X_n)$ to be invertible. If rank $< p \to$ infinite solutions, we pick the optimal β in a subset).

Ridge Regression:
What happens when $\Theta \subset \mathbb{R}^p$ with the optimization??