5th Lecture: Asymptotic (Large Sample) Properties of OLS estimator. 14/11/2022

↳ Main Reference: Chapter 5 in Greene.

• Recap:
↳ $E(\hat{\beta}|X) = \beta$, i.e. $\hat{\beta}$ is an unbiased estimator

↳ $E(s^2|X) = \sigma^2$, i.e. $s^2 = \frac{1}{n-k}\sum \hat{\varepsilon}_i^2$ unbiasc of σ^2

↳ $Var(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$

↳ GM Theorem: OLS is BLUE

↳ Extra Gaussian assumption: $\varepsilon|X \sim N(0, \sigma^2)$ allows statistical inference through Hypothesis Testing.

• Asymptotic Analysis: Useful when we are not sure about the exogeneity of the regressors (they may depend on y) and/or it is not reasonable to assume Gaussian errors.
However, we still assume a linear relationship between the response and the covariates.

↳ We make an alternative assumption:

The sample size is sufficiently large (or $n \to \infty$).

↳ We aim to approximate the (true) distribution of our estimators (rather than pre-assuming a form e.g. Gaussian)

- In particular we will answere the following questions:

↳ We have that OLS is $\hat{\theta}_n = \theta + (\underset{n\times k}{X'}\underset{n\times k}{X})^{-1}\underset{k\times n}{X'}\underset{n\times 1}{\varepsilon}$

and $\hat{\sigma}^2 = \dfrac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}$.

↳ We let $n \to \infty$ and we wish to find the limit of $\hat{\theta}_n$ ~~and~~ as well as its distribution.

- Preliminaries from Probability Theory

- Convergence in Probability:

Let $\{X_n\}$ be a seq. of r.v.'s then we say that $\{X_n\}$ converges in probability to a constant if for any $\varepsilon > 0$

$$\lim_{n \to \infty} \Pr(|X_n - c| > \varepsilon) = 0$$

Notation: $\operatorname{plim}_{n \to \infty} X_n = c$ or $\boxed{X_n \xrightarrow{p} c}$

- Mean Square Convergence

Let c fixed number s.t.

$$\lim_{n \to \infty} E\left[(X_n - c)^2 \right] = 0$$

then we say than $\{X_n\}$ converges in mean square to c and we write $X_n \xrightarrow{m.s.} c$

- Convergence in distribution

↳ Let $\{X_n\}$ ~~are~~ seq of random variables and

$\{F_n\}$ the corresponding seq of their cdfs.

If $F_n \xrightarrow[n\to\infty]{} F$ (cdf of X) or $|F_n(x) - F(x)| < \varepsilon \quad \forall \varepsilon > 0$

(Reminder: F_n and F are continuous functions)
not any randomness.

then we say than X_n converges in distr. to X

and we write $X_n \xrightarrow{d} X$.

Example:

$X_n \xrightarrow{d} N(0,1)$
implies that for $n\to\infty$ the sequence $\{X_n\}$
~~sequence~~ can be understood as a random
variable which follows $N(0,1)$

or for $n > N$ each element of
$\{X_n\}$ follows $N(0,1)$.

~~Preliminary~~ Remarks:

i) if $X_n \xrightarrow{P} X$ then $X_n \xrightarrow{d} X$

but the reverse is not necessarily true

ii) Each element of the sequence $\{X_n\}$ may follows a completely different distribution of the (limiting) distribution of X.

Example: $X_i \sim$ Binomial for $i = 1, \dots, n$

But for $n \to \infty$ Binomial can converge to a Gaussian distribution.

↳ Theoretical Exercises:

Proof that:

1) If $\lim_{n\to\infty} E X_n = c$ and $\lim_{n\to\infty} Var(X_n) = 0$

then $X_n \xrightarrow{m.s} c$

2) If $X_n \xrightarrow{m.s.} c \Rightarrow X_n \xrightarrow{P} c$

(6)

Proofs

1) We have that

$$E\left((X_n-c)^2\right) =$$

$$= E\left(\left[(X_n - EX_n) + (EX_n - c)\right]^2\right)$$

$$\overset{\text{check}}{=} E\left((X_n - EX_n)^2\right) + (EX_n - c)^2$$

$$= \underbrace{Var X_n}_{\xrightarrow{n\to\infty} 0} + (\underbrace{EX_n - c}_{\xrightarrow{n\to\infty} 0})^2$$

So $\lim E(X_n-c)^2 = 0$

i.e. $X_n \xrightarrow{m.s} c$

2) Reminder: Chebychev's inequality $P(|X-\mu|>k\sigma) \le \frac{1}{k^2}$

We have to show that $\lim_{n\to\infty} Pr(|X_n - c| \ge \varepsilon) = 0$.

For $\varepsilon > 0$ we have from Cheb. Ineq.

$$P(|X_n - c| \ge \varepsilon) \le \frac{1}{\varepsilon^2}$$

but $0 \le Pr(|X_n - c| \ge \varepsilon) \le \frac{1}{\varepsilon^2}$ $\Big\} \Rightarrow 0 \le P(|X_n - c| \ge \varepsilon) \le \frac{E(X_n-c)^2}{\varepsilon^2}$

and $E((X_n-c)^2) \ge 0$ grow the upper limit

Since the inequality holds for all n

we have that

$$0 \le \lim_{n\to\infty} P(|X_n - c| \ge \varepsilon) \le \lim_{n\to\infty} \frac{E((X_n-c)^2)}{\varepsilon^2} = 0$$

and thus $\lim_{n\to\infty} P(|X_n - c| \ge \varepsilon) = 0.$

4 Slutsky's Theorem:

If $\lim_{n\to\infty} P(|X_n - c| \ge \varepsilon) = 0 \quad \forall \varepsilon > 0$

then for any $g(\cdot)$ continuous

$$\lim_{n\to\infty} P(|g(X_n) - c| \ge \varepsilon) =$$

$$= g\left(\lim_{n\to\infty} P(|X_n - c| \ge \varepsilon)\right) = g(c)$$

* Remark Slutsky's Theorem says that limit passes through the function

(if we know $X_n \xrightarrow{P} c \Rightarrow g(x_n) \xrightarrow{P} g(c)$.

Examples:

- $X_n \xrightarrow{p} c \Rightarrow X_n^2 \xrightarrow{p} c^2$

- $X_n \xrightarrow{p} c$ and $Y_n \xrightarrow{p} d \Rightarrow X_n + Y_n \xrightarrow{p} c+d$

- $\underbrace{\hat{\beta}_n}_{\text{OLS}} \xrightarrow{p} \beta \Rightarrow \hat{\beta}_n^2 \xrightarrow{p} \beta^2$

- we also know $E\hat{\beta}_n = \beta$ but ~~[illegible]~~ $E\hat{\beta}_n^2 \neq \beta^2$

  and therefore this is useful as the best we can do

1) Law of Large Numbers

$$\underbrace{\bar{X}_n}_{\frac{1}{n}\sum_{i=1}^{n} X_i} \xrightarrow{p} \lim_{n\to\infty} E\bar{X}_n = \lim_{n\to\infty} E\left(\frac{\sum X_i}{n}\right) \quad (1)$$

* Very useful special case

Let $\{X_1, .., X_n\}$ random sample (no distribution assumption!)

s.t $EX_i = \mu$ and $Var X_i = \sigma^2$

then $\bar{X}_n \xrightarrow{p} \mu = EX_i$

4 Central Limit Theorem (CLT):

generic form: $\dfrac{\bar{X}_n - E\bar{X}_n}{\sqrt{Var\bar{X}_n}} \xrightarrow{d} N(0,1)$

~~Usefull~~

Usual Form:

If $X_1, \ldots, X_n$ random sample $\left(\begin{matrix} EX_i = \mu \\ Var X_i = \sigma^2 \end{matrix}\right)$ ↳ identically distributed

then $\dfrac{(\bar{X}_n - \mu)}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$

or $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0,\sigma^2)$

Remark: Notice that although

$X_n \to \mu \xrightarrow{n\to\infty} 0$

and $\sqrt{n} \xrightarrow{n\to\infty} \infty$

their product converges to something finite which is $N(0,\sigma^2)$. $\Rightarrow \sqrt{n}$ and $(\bar{X}_n - \mu)$ move at the same rate, this why we refer to the $\sqrt{n}$ - consistency of $\bar{X}_n$. or $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ for fixed large n.

Example: let $X_1, .., X_n$ random sample (μ, σ^2)

then for $S_n^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X}_n)^2}{n-1}$

we have that

$$\lim_{n \to \infty} Pr(|S_n^2 - \sigma^2| \geq \varepsilon) = 0$$

or $$p\lim_{n \to \infty} S_n^2 = \sigma^2$$

Proof: Find from Greene, or H/W or ask me later or at the end of the course if we find time

↳ Back to large sample properties for OLS

Let the extra (to Gauss Markov) assumptions

- $\{x_i, \varepsilon_i\}$, $i=1,..,n$ iid vectors

- $\text{plim}_{n\to\infty} \frac{X'X}{n} = Q$

$$= \text{plim}_{n\to\infty} \frac{\sum x_i x_i'}{n}$$

this is because $X'X = \sum_{i=1}^{n} \underset{k\times1}{x_i} \underset{1\times k}{x_i'}$

↳ discuss ~~see~~ this on p.14 for $X'\varepsilon = \sum_{i=1}^{n} x_i \varepsilon_i$

then $\hat{\beta}$ is a consistent estimator of β, i.e.

$$\text{plim}_{n\to\infty} \hat{\beta} = \beta.$$

$\boxed{\text{* Notice that from LLN definition } Q \text{ can be also written as } Q = \lim_{n\to\infty} E \frac{X'X}{n} = E(x_i x_i')}$

Proof of the Theorem:

We have that $\hat{\beta} = \beta + (X'X)^{-1} X'\varepsilon =$

$$= \beta + \left(\frac{X'X}{n}\right)^{-1} \frac{X'\varepsilon}{n}$$

Thus, $\text{plim}_{n\to\infty} \hat{\beta} = \beta + \text{plim}_{n\to\infty} \left(\frac{X'X}{n}\right)^{-1} \text{plim} \frac{X'\varepsilon}{n}$

$$= \beta + Q^{-1} \, \text{plim}_{n\to\infty} \frac{X'\varepsilon}{n}.$$

$(\cancel{\ldots}) = 1\times 1$

↳ reasonable sice $\underset{k\times n}{X'}\underset{n\times 1}{\varepsilon} = \boxed{k\times 1}$

- ① But $E(\cancel{\ldots}\underbrace{x_i\varepsilon_i}_{k\times 1\ \ 1\times 1}) =$

$$= E(E(x_i\varepsilon_i | X)) =$$

$$= E(x_i E(\varepsilon_i | X)) = \underset{k\times 1}{0}$$

and

$$Var(\underset{k\times 1}{x_i}\underset{1\times 1}{\varepsilon_i}) = E(\underset{k\times 1}{x_i}\underset{1\times 1}{\varepsilon_i}(\underset{k\times 1}{x_i}\underset{1\times 1}{\varepsilon_i})') =$$

$$= E(\underset{k\times 1}{x_i}\underset{1\times k}{x_i'}\underset{1\times 1}{\varepsilon_i^2})$$

$$= E(E(x_i x_i' \varepsilon_i^2 | x_i))$$

$$= E(x_i x_i' E(\varepsilon_i^2 | x_i)) =$$

$$= \sigma^2 E(x_i x_i') =$$

↳ vector with k elements $(k\times 1)$

$$= \sigma^2 \underset{k\times k}{Q_i} \quad \left(\text{remember } Q = \frac{X'X}{n}\right) \quad (1)$$

② Define the random variable

$$\bar{W} = \frac{\sum_{i=1}^{n} x_i\varepsilon_i}{n}, \quad \underset{k\times 1}{W_i} = x_i\varepsilon_i$$

then $E\bar{W} = E(E(x_i\varepsilon_i | x_i)) = E(x_i E(\varepsilon_i | x_i)) = 0$

and $Var\,\bar{W} \overset{\text{indep}}{=} \sum_{i=1}^{n} Var\,\frac{x_i\varepsilon_i}{n^2}$

$$\overset{(1)}{=} \sigma^2 \frac{\sum_{i=1}^{n} Q_i}{n^2} = \frac{\sigma^2}{n} E\underbrace{\left(\frac{X'X}{n}\right)}_{k\times k} \xrightarrow[n\to\infty]{} 0 \quad (2)$$

So for $\bar{W}$ we have

$$\left.\begin{array}{l}\lim\limits_{n\to\infty} \bar{W} = 0 \\ \text{and } \lim\limits_{n\to\infty} \mathrm{Var}\, \bar{W} = 0\end{array}\right\} \Rightarrow \bar{W} \xrightarrow{m.s.} 0 \Rightarrow$$

$$\overset{\parallel}{\left(\frac{\sum x_i \varepsilon_i}{n}\right)}$$

$$\Rightarrow \frac{\sum \overset{k\times 1}{x_i} \varepsilon_i}{n} \xrightarrow{p} 0$$

From p. 11 we have that

$$\underset{n\to\infty}{\mathrm{plim}}\, \hat{\beta} = \beta + Q^{-1}\, \underset{n\to\infty}{\mathrm{plim}} \frac{X'\varepsilon}{n}$$

$$\parallel \quad \text{vector } k\times 1$$

$$\underset{n\to\infty}{\mathrm{plim}} \frac{\sum x_i \overset{1\times 1}{\varepsilon_i}}{n} \xrightarrow{p} 0$$

and thus $\underset{n\to\infty}{\mathrm{plim}}\, \hat{\beta} \xrightarrow{p} \beta$

→ Asymptotic Normality of the OLS $\hat{\beta}$ (14)

Write $\sqrt{n}(\hat{\beta}-\beta) \overset{(X'X)^{-1}X'\varepsilon}{=} \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\varepsilon}{\sqrt{n}}\right)$

$\left(\frac{X'X}{n}\right)^{-1} \xrightarrow{p} Q^{-1}$

For the other term we have

$$\frac{\overset{k\times 1}{X'\varepsilon}}{\sqrt{n}} = \sqrt{n}\left(\frac{\sum_{i=1}^{n} \overset{w_i}{x_i\varepsilon_i}}{n} - 0\right) =$$

$x_i\varepsilon_i$: $k\times 1$

$$= \sqrt{n}\left(\bar{w} - E\bar{w}\right)$$

and $\operatorname{Var}(\sqrt{n}\,\bar{w}) = n\operatorname{Var}\bar{w} =$

$$\overset{\text{p. 12 other cond}}{=} \sigma^2 E\left(\frac{X'X}{n}\right)$$

$$= \sigma^2 Q$$

** $X'\varepsilon = \sum \overset{k\times 1}{x_i}\overset{1\times 1}{\varepsilon_i}$

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ \vdots & \vdots & \vdots \\ x_{k1} & x_{k2} & x_{k3} \end{pmatrix}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \sum x_{1i}\varepsilon_i \\ \sum x_{2i}\varepsilon_i \\ \sum x_{3i}\varepsilon_i \end{pmatrix} = \sum \underset{k\times 1}{x_i}\,\varepsilon_i$$

↳ we first multiply x_{1i}, x_{2i}, x_{3i} with ε_i and then take the sum

Then from the (Lindenberg-Feller) CLT → Greene p. 1079 D.19

↳ generalization of CLT for not identically distributed r.v.

↳ i.e. unequal means and variances.

$$\sqrt{n}(\bar{X}_n - \bar{\mu}_n) \xrightarrow{d} N(0, \bar{\sigma}^2)$$

$\bar{\mu}_n = \frac{\sum \mu_i}{n}$, $\bar{\sigma}_n^2 = \frac{\sum \sigma_i^2}{n}$, $\bar{\sigma}^2 = \lim_{n\to\infty} \bar{\sigma}_n^2$

we have that

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} Q^{-1}N(0, \sigma^2 Q) \equiv N(0, \sigma^2 Q^{-1})$$

* Note that σ^2 and Q are population quantities

↳ $Q = \lim_{n\to\infty} E\frac{X'X}{n}$

13

So for $\bar{W}$ we have

$$\left.\begin{array}{l} \lim_{n\to\infty} \bar{W} = 0 \\ \text{and } \lim_{n\to\infty} \operatorname{Var} \bar{W} = 0 \end{array}\right\} \Rightarrow \bar{W} \xrightarrow{m.s.} 0 \Rightarrow$$

$$\overset{\parallel}{\left(\frac{\sum x_i \varepsilon_i}{n}\right)}$$

$$\Rightarrow \frac{\sum \overset{k\times 1}{x_i} \varepsilon_i}{n} \xrightarrow{p} 0$$

From p. 11 we have that

$$\operatorname{plim}_{n\to\infty} \hat{\beta} = \beta + Q^{-1} \operatorname{plim}_{n\to\infty} \frac{X'\varepsilon}{n}$$

$$\parallel$$

$$\operatorname{plim}_{n\to\infty} \frac{\sum \overset{\text{vector } k\times 1}{x_i} \overset{1\times 1}{\varepsilon_i}}{n} \xrightarrow{p} 0$$

and thus $\operatorname{plim}_{n\to\infty} \hat{\beta} \to \beta$

<u>Remark 1</u>: The asymptotic normality of the OLS $\hat{\beta}$ does not depend on the normality of the error term ε. It is just a consequence of the CLT.

<u>Remark 2</u>: Let $\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}$ the estimator of σ^2 and $\frac{X'X}{n}$ the estimator of Q then a vector form of the Lidenberg-Feller CLT gives that

$$\hat{\beta} \sim N(\beta, \hat{\sigma}^2 (X'X)^{-1}) \quad (*)$$

<u>Remark 3</u>: Notice the similarity of (*) with the finite sample formula but σ^2 has replaced by $\hat{\sigma}^2$ and normality still holds whereas normality is again without any distributional assumption for the errors ε.

- Consistency of $\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}$

↳ We can show that

$$\underset{n\to\infty}{\text{plim}}\, \hat{\sigma}^2 = \sigma^2.$$

(Proof: at the end of the course if we find time)

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k} = \frac{\varepsilon' M \varepsilon}{n-k} = \frac{\varepsilon'\left(I - X(X'X)^{-1}X'\right)\varepsilon}{n-k} =$$

$$= \frac{\varepsilon'\varepsilon - \varepsilon' X(X'X)^{-1}X'\varepsilon}{n-k}$$

$$= \underbrace{\frac{n}{n-k}}_{\to 1}\left(\frac{\varepsilon'\varepsilon}{n} - \underbrace{\frac{\varepsilon' X}{n}}_{\xrightarrow{p} 0}\underbrace{\left(\frac{X'X}{n}\right)^{-1}}_{\xrightarrow{p} Q^{-1}}\underbrace{\frac{X'\varepsilon}{n}}_{0_P}\right)$$

from previous (p. 13)

Thus $\hat{\sigma}^2 \xrightarrow{p} \frac{\varepsilon'\varepsilon}{n} = \sigma^2$

Since $\frac{\varepsilon'\varepsilon}{n} = \frac{\sum_{i=1}^{n}\varepsilon_i^2}{n} \xrightarrow[\text{LLN}]{p} \frac{E\sum_{i=1}^{n}\varepsilon_i^2}{n} =$

$$= \frac{\sum_{i=1}^{n} E\left(\overbrace{E(\varepsilon_i^2 \mid x_i)}^{\sigma^2}\right)}{n} = \sigma^2$$

↳ Asymptotic Normality of t-statistic:

~~We have seen that~~ CLT implies that

$$Z_j = \frac{\sqrt{n}\,(\hat{\beta}_j - \beta_j)}{\sqrt{\sigma^2 (X'X)^{-1}_{jj}}} = \frac{\sqrt{n}\,(\hat{\beta}_j - \beta_j)}{\sqrt{\sigma^2 Q^{-1}_{jj}}} \xrightarrow{d} N(0,1)$$

Then we can also show that

$$t_j = \frac{\sqrt{n}(\hat{\beta}_j - \beta_j)}{\sqrt{\hat{\sigma}^2 (X'X)^{-1}_{jj}}} \xrightarrow{d} N(0,1)$$

↳ Notice the difference with finite sample where we had the t-distribution

(Proof at the end of the course).