Advanced Econometrics

Lecture 7: Large Sample Theory (Hansen ch. 6)

Instructor: Ma, Jun

Renmin University of China

November 15, 2021

Why we need the large sample theory

- We have shown that the OLS estimator $\hat{\boldsymbol{\beta}}$ has some desirable properties:
  - $\hat{\boldsymbol{\beta}}$ is unbiased if the errors are strongly exogenous: $\mathbb{E}(\boldsymbol{e} \mid \boldsymbol{X}) = \mathbf{0}$.
  - If in addition the errors are homoskedastic then $\hat{\boldsymbol{V}}^0_{\hat{\boldsymbol{\beta}}} = (\boldsymbol{X}'\boldsymbol{X})^{-1} s^2$ is an unbiased estimator of the conditional variance of the OLS estimator $\hat{\boldsymbol{\beta}}$.
  - If in addition the errors are normally distributed (given $\boldsymbol{X}$) then the t statistic has a t distribution which can be used for hypotheses testing.

- If the errors are only weakly exogenous: $\mathbb{E}\left(\boldsymbol{X}_i e_i\right)=0$ the OLS estimator is in general biased.
- If the errors are heteroskedastic:$\mathbb{E}\left(e_i^2 \mid \boldsymbol{X}_i\right)=h\left(\boldsymbol{X}_i\right)$, the "usual" variance formula is invalid; we also do not have an unbiased estimator for the variance in this case.
- If the errors are not normally distributed conditional on $\boldsymbol{X}$ then t and F statistics do not have t and F distributions under the null hypothesis.
- The asymptotic or large sample theory allows us to derive approximate properties and distributions of estimators and test statistics by assuming that the sample size n is very large.

- We will argue that even when the errors are not normally distributed, the OLS estimator has an approximately normal distribution in large samples, provided that some additional conditions hold.
- This property is used for hypothesis testing: in large samples, the t statistic has a standard normal distribution and the F statistic has a χ^2 distribution (approximately).

Limits and convergence concepts

- The concept of convergence cannot be applied in a straightforward way to sequences of random variables. This is so because a random variable is a function from the sample space Ω to the real line. The solution is to consider convergence of a non-random sequence derived from the random one.
- Let $\{X_n : n = 1, 2, \ldots\}$ be a sequence of random variables. Let X be random or non-random.
- We will consider non-random sequences with the following typical elements: 1. $\mathbb{E}\,|X_n - X|^r$; 2. $\Pr\left(|X_n - X| > \varepsilon\right)$ for some $\varepsilon > 0$.
  - Convergence in r-th mean. X_n converges to X in r-th mean if $\mathbb{E}\,|X_n - X|^r \to 0$ as $n \to \infty$.
  - Convergence in probability. X_n converges in probability to X if for all $\varepsilon > 0$, $\Pr\left(|X_n - X| \geq \varepsilon\right) \to 0$ as $n \to \infty$. It is denoted as $X_n \to_p X$ or $\operatorname{plim}_{n\to\infty} X_n = X$.

- Convergence in r-th mean implies convergence in probability.
- (Markov's Inequality) Let X be a random variable. For $\varepsilon > 0$ and $r > 0$,

$$\Pr(|X| \geq \varepsilon) \leq \mathbb{E}|X|^r / \varepsilon^r.$$

- Suppose that X_n converges to X in r-th mean, $\mathbb{E}|X_n - X|^r \to 0$. Then,

$$\begin{aligned} \Pr(|X_n - X| \geq \varepsilon) &\leq \mathbb{E}|X_n - X|^r / \varepsilon^r \\ &\to 0. \end{aligned}$$

Rules for probability limits

Suppose that $X_n \to_p a$ and $Y_n \to_p b$, where a and b are some finite constants. Let c be another constant.

- $cX_n \to_p ca$.
- $X_n + Y_n \to_p a + b$.
- $X_n Y_n \to_p ab$.
- $X_n / Y_n \to_p a/b$, provided that $b \neq 0$.
- If $0 \leq X_n \leq Y_n$ and $Y_n \to_p 0$, then $X_n \to_p 0$.
- $X_n \to_p 0$ if and only if $|X_n| \to_p 0$.

Continuous mapping theorem (CMT)

- Suppose that $X_n \to_p c$, a constant, and let $h(\cdot)$ be a continuous function at c. Then, $h(X_n) \to_p h(c)$.
- suppose that $\widehat{\beta}_n \to_p \beta$. Then $\widehat{\beta}_n^2 \to_p \beta^2$, and $1/\widehat{\beta}_n \to_p 1/\beta$, provided $\beta \neq 0$.

Convergence of random vectors

- The random vectors/matrices converge in probability if their elements converge in probability.
- Consider the vector case. Let $\{\boldsymbol{X}_n : n = 1, 2, \ldots\}$ be a sequence of random k-vectors. $\boldsymbol{X}_n - \boldsymbol{X} \to_p 0$ element-by-element, where $\boldsymbol{X}$ is a possibly random k-vector, if and only if $\|\boldsymbol{X}_n - \boldsymbol{X}\| \to_p 0$, where $\|\cdot\|$ denotes the Euclidean norm.
- The rules for manipulation of probability limits in the vector/matrix case are similar to those in the scalar case.
- The CMT is valid in vector/matrix case as well.

Weak law of large numbers

- Let $X_1, \dots X_n$ be a sample of iid random variables such that $\mathbb{E}\,|X_1| < \infty$. Then, $n^{-1}\sum_{i=1}^{n} X_i \to_p \mathbb{E}X_1$ as $n \to \infty$.
- Due to iid assumption, we have that $\mathbb{E}X_i = \mathbb{E}X_1$ for all $i = 1, \dots, n$.

Convergence in distribution

- Let $\{X_n : n = 1, 2, \ldots\}$ be a sequence of random variables.
- Let $F_n(x)$ denote the marginal CDF of X_n, i.e. $F_n(x) = \Pr(X_n \leq x)$. Let $F(x)$ be another CDF.
- We say that X_n converges in distribution if $F_n(x) \to F(x)$ for all x where $F(x)$ is continuous.
- In this case, we write $X_n \to_d X$, where X is any random variable with the distribution function $F(x)$.
- Note that while we say that X_n converges to X, the convergence in distribution is not convergence of random variables, but of the distribution functions.

- The extension to the vector case is straightforward. Let $\boldsymbol{X}_n$ and $\boldsymbol{X}$ be two random k-vectors.
- We say that $\boldsymbol{X}_n \to_d \boldsymbol{X}$ if the joint CDF of $\boldsymbol{X}_n$ converges to that of $\boldsymbol{X}$ at all continuity points, i.e.

$$\begin{aligned} F_n(x_1, \ldots, x_k) &= \Pr(X_{n,1} \leq x_1, \ldots, X_{n,k} \leq x_k) \\ &\to \Pr(X_1 \leq x_1, \ldots, X_k \leq x_k) \\ &= F(x_1, \ldots, x_k), \end{aligned}$$

  for all points $(x_1, \ldots, x_k)$ where F is continuous.
- In this case, we say that the elements of $\boldsymbol{X}_n$, $X_{n,1}, \ldots X_{n,k}$, jointly converge in distribution to $X_1, \ldots X_k$, the elements of $\boldsymbol{X}$.

Rules of convergence in distribution

- (Cramer Convergence Theorem) Suppose that $X_n \to_d X$ and $Y_n \to_p c$. Then,
  - $X_n + Y_n \to_d X + c$.
  - $Y_n X_n \to_d cX$.
  - $X_n / Y_n \to_d X/c$, provided that $c \neq 0$.
- If $X_n \to_p X$, then $X_n \to_d X$. Converse is not true with one exception: If $X_n \to_d c$, a constant, then $X_n \to_p c$.
- If $X_n - Y_n \to_p 0$, and $Y_n \to_d Y$, then $X_n \to_d Y$.

Continuous mapping theorem

- Suppose that $X_n \to_d X$, and let $h(\cdot)$ be a function continuous on a set $\mathcal{X}$ such that $\Pr(X \in \mathcal{X}) = 1$. Then, $h(X_n) \to_d h(X)$.
- Examples:
  - Suppose that $X_n \to_d X$. Then $X_n^2 \to_d X^2$. For example, if $X_n \to_d N(0,1)$, then $X_n^2 \to_d \chi_1^2$.
  - Suppose that $(X_n, Y_n) \to_d (X, Y)$ (joint convergence in distribution), and set $h(x,y) = x$. Then $X_n \to_d X$. Set $h(x,y) = x^2 + y^2$. Then $X_n^2 + Y_n^2 \to_d X^2 + Y^2$. For example, if $(X_n, Y_n) \to_d N(0, I_2)$, then $X_n^2 + Y_n^2 \to_d \chi_2^2$.
- Note that contrary to convergence in probability, $X_n \to_d X$ and $Y_n \to_d Y$ does not imply that, for example, $X_n + Y_n \to_d X + Y$, unless a joint convergence result holds.

The central limit theorem

- Let $X_1, \ldots, X_n$ be a sample of iid random variables such that $\mathbb{E}X_1 = 0$ and $0 < \mathbb{E}X_1^2 < \infty$. Then, as $n \to \infty$, $n^{-1/2} \sum_{i=1}^n X_i \to_d N(0, \mathbb{E}X_1^2)$.
- Let $X_1, \ldots X_n$ be a sample of iid random variables with $\mathbb{E}X_1 = \mu$ and $\text{Var}\,(X_1) = \sigma^2 < \infty$. Define

$$\overline{X}_n = n^{-1} \sum_{i=1}^n X_i .$$

- Consider $n^{-1/2} \sum_{i=1}^n (X_i - \mu)$. We have that $(X_1 - \mu), \ldots, (X_n - \mu)$ are iid with the mean $\mathbb{E}\,(X_1 - \mu) = 0$, and the variance $\mathbb{E}\,(X_1 - \mu)^2 = \sigma^2 < \infty$. Therefore, by the CLT,

$$\begin{aligned} n^{1/2}\left(\overline{X}_n - \mu\right) &= n^{-1/2} \sum_{i=1}^n (X_i - \mu) \\ &\to_d N\left(0, \sigma^2\right). \end{aligned}$$

- In practice, we use convergence in distribution as an approximation. Let $\overset{a}{\sim}$ denote "approximately in large samples".
- Informally, one can say that $n^{1/2}\left(\overline{X}_n - \mu\right) \overset{a}{\sim} N\left(0, \sigma^2\right)$ or $\overline{X}_n \overset{a}{\sim} N\left(\mu, \sigma^2/n\right)$.
- Note that under the normality assumption for X_i's, the above result is obtained exactly for any sample size n.

Cramer-Wold device and multivariate CLT

- Let $\boldsymbol{X}_n$ be a random k-vector. Then, $\boldsymbol{X}_n \to_d \boldsymbol{X}$ if and only if $\boldsymbol{\lambda}'\boldsymbol{X}_n \to_d \boldsymbol{\lambda}'\boldsymbol{X}$ for all non-zero $\boldsymbol{\lambda} \in \mathbb{R}^k$.
- Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a sample of iid random k-vectors such that $\mathbb{E}\boldsymbol{X}_1 = \boldsymbol{0}$ (denote $\boldsymbol{X}_i = (X_{i,1}, ..., X_{i,k})'$) and $\mathbb{E}X_{1,j}^2 < \infty$ for all $j = 1, \ldots, k$, and $\mathbb{E}\left(\boldsymbol{X}_1\boldsymbol{X}_1'\right)$ is positive definite. Then, $n^{-1/2}\sum_{i=1}^n \boldsymbol{X}_i \to_d N(0, \mathbb{E}\left(\boldsymbol{X}_1\boldsymbol{X}_1'\right))$.

Delta method

- The delta method is used to derive the asymptotic distribution of the nonlinear functions of estimators. For example, $\overline{X}_n \to_p \mathbb{E}X_1 = \mu$. It follows from the CMT that $h\left(\overline{X}_n\right) \to_p h(\mu)$. The delta method provides approximation for the distribution of $h\left(\overline{X}_n\right)$.
- We have that $n^{1/2}\left(\overline{X}_n - \mu\right) \to_d N\left(0, \sigma^2\right)$. Suppose that $\mu \neq 0$. Then, by the delta method,

$$\begin{aligned} n^{1/2}\left(\frac{1}{\overline{X}_n} - \frac{1}{\mu}\right) &\to_d -\frac{1}{\mu^2} N\left(0, \sigma^2\right) \\ &= N\left(0, \frac{\sigma^2}{\mu^4}\right). \end{aligned}$$

- Let $\widehat{\boldsymbol{\theta}}_n$ be a random k-vector, and suppose that $n^{1/2}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \to_d \boldsymbol{Y}$ as $n \to \infty$, where $\boldsymbol{\theta}$ is a k-vector of constants ($\boldsymbol{\theta} = (\theta_1, ..., \theta_k)'$), and $\boldsymbol{Y}$ is a random k-vector. Let $\boldsymbol{h} : \mathbb{R}^k \to \mathbb{R}^m$ be a function continuously differentiable on some open neighborhood of $\boldsymbol{\theta}$. Equivalently, we can denote $\boldsymbol{h} = (h_1, ..., h_m)'$, where $h_j : \mathbb{R}^k \to \mathbb{R}$, $j = 1, ..., m$. Then, $n^{1/2}\left(\boldsymbol{h}\left(\widehat{\boldsymbol{\theta}}_n\right) - \boldsymbol{h}(\boldsymbol{\theta})\right) \to_d \frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} Y$, where

$$\frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \begin{pmatrix} \frac{\partial h_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\ \vdots \\ \frac{\partial h_m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \end{pmatrix} = \begin{pmatrix} \frac{\partial h_1(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial h_1(\boldsymbol{\theta})}{\partial \theta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial h_m(\boldsymbol{\theta})}{\partial \theta_k} \end{pmatrix}.$$