Introductory Econometrics

Lecture 15: Large sample results: Consistency

Instructor: Ma, Jun

Renmin University of China

November 16, 2022

Why we need the large sample theory

- We have shown that the OLS estimator $\hat{\beta}$ has some desirable properties:
  - $\hat{\beta}$ is unbiased if the errors are strongly exogenous: $\mathrm{E}\left[U \mid X\right] = 0$.
  - If in addition the errors are homoskedastic then $\widehat{\mathrm{Var}}\left(\hat{\beta}\right) = s^2 / \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ is an unbiased estimator of the conditional variance of the OLS estimator $\hat{\beta}$.
  - If in addition the errors are normally distributed (given X) then $T = \left(\hat{\beta} - \beta\right) / \sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}$ has a t distribution which can be used for hypotheses testing.

- If the errors are only weakly exogenous:

$$\mathrm{E}\left[X_i U_i\right] = 0,$$

  the OLS estimator is in general biased.
- If the errors are heteroskedastic:

$$\mathrm{E}\left[U_i^2 \mid X_i\right] = h\left(X_i\right),$$

  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 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.

Convergence in probability and LLN

- Let θ_n be a sequence of random variables indexed by the sample size n. We say that θ_n converges in probability if

$$\lim_{n\to\infty} \Pr\left[|\theta_n - \theta| \geq \varepsilon\right] = 0 \text{ for all } \varepsilon > 0.$$

- We denote this as $\theta_n \to_p \theta$ or $\operatorname{plim} \theta_n = \theta$.
- An example of convergence in probability is a Law of Large Numbers (LLN):
  Let $X_1, X_2, \ldots, X_n$ be a random sample such that $\mathrm{E}\left[X_i\right] = \mu$ for all $i = 1, \ldots, n$, and define $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
  Then, under certain conditions,

$$\bar{X}_n \to_p \mu.$$

LLN

- Let $X_1, \ldots, X_n$ be a sample of independent identically distributed (iid) random variables. Let $\mathrm{E}[X_i] = \mu$. If $\mathrm{Var}[X_i] = \sigma^2 < \infty$ then
$$\bar{X}_n \to_p \mu.$$
- In fact when the data are iid, the LLN holds if
$$\mathrm{E}[|X_i|] < \infty,$$
but we prove the result under a stronger assumption that $\mathrm{Var}[X_i] < \infty$.

Markov's inequality

- Markov's inequality. Let W be a random variable. For $\varepsilon > 0$ and $r > 0$,

$$\Pr[|W| \geq \varepsilon] \leq \frac{\mathrm{E}[|W|^r]}{\varepsilon^r}.$$

- With $r = 2$, we have Chebyshev's inequality. Suppose that $\mathrm{E}[X] = \mu$. Take $W = X - \mu$ and apply Markov's inequality with $r = 2$. For $\varepsilon > 0$,

$$\begin{aligned} \Pr[|X - \mu| \geq \varepsilon] &\leq \frac{\mathrm{E}\left[(X - \mu)^2\right]}{\varepsilon^2} \\ &= \frac{\mathrm{Var}[X]}{\varepsilon^2}. \end{aligned}$$

- Probability of observing an outlier (a large deviation of X from its mean μ) can be bounded by the variance.

Proof of the LLN

Fix $\varepsilon > 0$ and apply Markov's inequality with $r = 2$:

$$\Pr\left[\left|\bar{X}_n - \mu\right| \geq \varepsilon\right] = \Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right]$$

$$= \Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} (X_i - \mu)\right| \geq \epsilon\right] \leq \frac{\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} (X_i - \mu)\right)^2\right]}{\epsilon^2}$$

$$= \frac{1}{n^2\epsilon^2}\left(\sum_{i=1}^{n} \mathrm{E}\left[(X_i - \mu)^2\right] + \sum_{i=1}^{n}\sum_{j\neq i} \mathrm{E}\left[(X_i - \mu)\left(X_j - \mu\right)\right]\right)$$

$$= \frac{1}{n^2\epsilon^2}\left(\sum_{i=1}^{n} \mathrm{Var}\left[X_i\right] + \sum_{i=1}^{n}\sum_{j\neq i} \mathrm{Cov}\left[X_i, X_j\right]\right)$$

$$= \frac{n\sigma^2}{n^2\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \text{ as } n \to \infty \text{ for all } \epsilon > 0.$$

Averaging and variance reduction

- Let $X_1, \ldots, X_n$ be a sample and suppose that

$$\begin{aligned} \mathrm{E}[X_i] &= \mu \text{ for all } i = 1, \ldots, n, \\ \mathrm{Var}[X_i] &= \sigma^2 \text{ for all } i = 1, \ldots, n, \\ \mathrm{Cov}\left[X_i, X_j\right] &= 0 \text{ for all } j \neq i. \end{aligned}$$

- Consider the mean of the average:

$$\begin{aligned} \mathrm{E}\left[\bar{X}_n\right] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[X_i] \\ &= \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu. \end{aligned}$$

- Consider the variance of the average:

$$\begin{aligned}
\mathrm{Var}\left[\bar{X}_n\right] &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{Var}\left[X_i\right] + \sum_{i=1}^{n}\sum_{j\neq i}\mathrm{Cov}\left[X_i, X_j\right]\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n}\sigma^2 + \sum_{i=1}^{n}\sum_{j\neq i} 0\right) \\
&= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

- The variance of the average approaches zero as $n \to \infty$ if the observations are uncorrelated.

Convergence in probability: properties

- Slutsky's Lemma. Suppose that $\theta_n \to_p \theta$, and let g be a function continuous at θ. Then,

$$g(\theta_n) \to_p g(\theta).$$

  - If $\theta_n \to_p \theta$, then $\theta_n^2 \to_p \theta^2$.
  - If $\theta_n \to_p \theta$ and $\theta \neq 0$, then $1/\theta_n \to_p 1/\theta$.
- Suppose that $\theta_n \to_p \theta$ and $\lambda_n \to_p \lambda$. Then,
  - $\theta_n + \lambda_n \to_p \theta + \lambda$.
  - $\theta_n \lambda_n \to_p \theta\lambda$.
  - $\theta_n / \lambda_n \to_p \theta/\lambda$ provided that $\lambda \neq 0$.

Consistency

- Let $\hat{\beta}_n$ be an estimator of β based on a sample of size n.
- We say that $\hat{\beta}_n$ is a consistent estimator of β if as $n \to \infty$,

$$\hat{\beta}_n \to_p \beta.$$

- Consistency means that the probability of the event that the distance between $\hat{\beta}_n$ and β exceeds $\varepsilon > 0$ can be made arbitrary small by increasing the sample size.

Consistency of OLS

- Suppose that:
  1. The data $\{(Y_i, X_i) : i = 1, \ldots, n\}$ are iid.
  2. $Y_i = \beta_0 + \beta_1 X_i + U_i$, where $\mathrm{E}[U_i] = 0$.
  3. $\mathrm{E}[X_i U_i] = 0$.
  4. $0 < \mathrm{Var}[X_i] < \infty$.
- Let $\hat{\beta}_{0,n}$ and $\hat{\beta}_{1,n}$ be the OLS estimators of β_0 and β_1 respectively based on a sample of size n. Under Assumptions 1-4,

$$\hat{\beta}_{0,n} \to_p \beta_0,$$
$$\hat{\beta}_{1,n} \to_p \beta_1.$$

- The key identifying assumption is Assumption 3: $\mathrm{Cov}[X_i, U_i] = 0$.

Proof of consistency

- Write

$$\hat{\beta}_{1,n} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)Y_i}{\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2} \quad = \quad \beta_1 + \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)U_i}{\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$$

$$= \quad \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)U_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}.$$

- We will show that

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)U_i \to_p 0,$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 \to_p \operatorname{Var}\left[X_i\right],$$

- Since $\operatorname{Var}\left[X_i\right] \neq 0$,

$$\hat{\beta}_{1,n} = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)U_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2} \to_p \beta_1 + \frac{0}{\operatorname{Var}\left[X_i\right]} = \beta_1.$$

$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)U_i \rightarrow_p 0$

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)U_i = \frac{1}{n}\sum_{i=1}^{n}X_iU_i - \bar{X}_n\left(\frac{1}{n}\sum_{i=1}^{n}U_i\right).$$

By the LLN,

$$\frac{1}{n}\sum_{i=1}^{n}X_iU_i \rightarrow_p \mathrm{E}\left[X_iU_i\right] = 0,$$

$$\bar{X}_n \rightarrow_p \mathrm{E}\left[X_i\right],$$

$$\frac{1}{n}\sum_{i=1}^{n}U_i \rightarrow_p \mathrm{E}\left[U_i\right] = 0.$$

Hence,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)U_i = \frac{1}{n}\sum_{i=1}^{n}X_iU_i - \bar{X}_n\left(\frac{1}{n}\sum_{i=1}^{n}U_i\right) &\rightarrow_p 0 - \mathrm{E}\left[X_i\right]\cdot 0 \\ &= 0.\end{aligned}$$

$\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}_n\right)^2 \to_p \mathrm{Var}\left[X_i\right]$

- First,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}_n\right)^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2-2\bar{X}_nX_i+\bar{X}_n^2\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}X_i^2-2\bar{X}_n\frac{1}{n}\sum_{i=1}^{n}X_i+\bar{X}_n^2\\ &= \frac{1}{n}\sum_{i=1}^{n}X_i^2-2\bar{X}_n\bar{X}_n+\bar{X}_n^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}_n^2.\end{aligned}$$

- By the LLN, $\frac{1}{n}\sum_{i=1}^{n}X_i^2 \to_p \mathrm{E}\left[X_i^2\right]$ and $\bar{X}_n \to_p \mathrm{E}\left[X_i\right].$
- By Slutsky's Lemma, $\bar{X}_n^2 \to_p \left(\mathrm{E}\left[X_i\right]\right)^2$.
- Thus,

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}_n\right)^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}_n^2 \to_p \mathrm{E}\left[X_i^2\right]-\left(\mathrm{E}\left[X_i\right]\right)^2=\mathrm{Var}\left[X_i\right].$$

Multiple regression

- Under similar conditions to 1-4, one can establish consistency of OLS for the multiple linear regression model:

  $$Y_i = \beta_0 + \beta_1 X_{1,i} + \ldots + \beta_k X_{k,i} + U_i,$$

  where $\mathrm{E}\left[U_i\right] = 0$.
- The key assumption is that the errors and regressors are uncorrelated:

  $$\mathrm{E}\left[X_{1,i} U_i\right] = \ldots = \mathrm{E}\left[X_{k,i} U_i\right] = 0.$$

Omitted variables and the inconsistency of OLS

- Suppose that the true model has two regressors:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + U_i, \\ \mathrm{E}\left[X_{1,i} U_i\right] &= \mathrm{E}\left[X_{2,i} U_i\right] = 0. \end{aligned}$$

- Suppose that the econometrician includes only X_1 in the regression when estimating β_1:

$$\begin{aligned} \tilde{\beta}_{1,n} &= \frac{\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right) Y_i}{\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)^2} \\ &= \frac{\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)\left(\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + U_i\right)}{\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)^2} \\ &= \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right) X_{2,i}}{\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)^2} + \frac{\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right) U_i}{\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)^2}. \end{aligned}$$

$$\tilde{\beta}_{1,n} = \beta_1 + \beta_2 \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right) X_{2,i}}{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)^2} + \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right) U_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)^2}.$$

- As before,

$$\begin{aligned}
\frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right) U_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)^2} =& \frac{\frac{1}{n}\sum_{i=1}^{n} X_{1,i}U_i - \bar{X}_{1,n}\bar{U}_n}{\frac{1}{n}\sum_{i=1}^{n} X_{1,i}^2 - \bar{X}_{1,n}} \\
\to_p& \frac{0}{\mathrm{E}\left[X_{1,i}^2\right] - \left(\mathrm{E}\left[X_{1,i}\right]\right)^2} \\
=& \frac{0}{\mathrm{Var}\left[X_{1,i}\right]} = 0.
\end{aligned}$$

$$\tilde{\beta}_{1,n} = \beta_1 + \beta_2 \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)X_{2,i}}{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)^2} + \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)U_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)^2}.$$

- However,

$$\begin{aligned}
\frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)X_{2,i}}{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)^2} &= \frac{\frac{1}{n}\sum_{i=1}^{n} X_{1,i}X_{2,i} - \bar{X}_{1,n}\bar{X}_{2,n}}{\frac{1}{n}\sum_{i=1}^{n} X_{1,i}^2 - \bar{X}_{1,n}^2} \\
&\to_p \frac{\mathrm{E}\left[X_{1,i}X_{2,i}\right] - \left(\mathrm{E}\left[X_{1,i}\right]\right)\left(\mathrm{E}\left[X_{2,i}\right]\right)}{\mathrm{E}\left[X_{1,i}^2\right] - \left(\mathrm{E}\left[X_{1,i}\right]\right)^2} \\
&= \frac{\mathrm{Cov}\left[X_{1,i}, X_{2,i}\right]}{\mathrm{Var}\left[X_{1,i}\right]}.
\end{aligned}$$

- We have,

$$
\begin{aligned}
\tilde{\beta}_{1,n} &= \beta_1 + \beta_2 \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right) X_{2,i}}{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)^2} + \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right) U_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i} - \bar{X}_{1,n}\right)^2} \\
&\to_p \beta_1 + \beta_2 \frac{\mathrm{Cov}\left[X_{1,i}, X_{2,i}\right]}{\mathrm{Var}\left[X_{1,i}\right]} + \frac{0}{\mathrm{Var}\left[X_{1,i}\right]} \\
&= \beta_1 + \beta_2 \frac{\mathrm{Cov}\left[X_{1,i}, X_{2,i}\right]}{\mathrm{Var}\left[X_{1,i}\right]}.
\end{aligned}
$$

- Thus, $\tilde{\beta}_{1,n}$ is inconsistent unless:
  1. $\beta_2 = 0$ (the model is correctly specified).
  2. $\mathrm{Cov}\left[X_{1,i}, X_{2,i}\right] = 0$ (the omitted variable is uncorrelated with the included regressor).

- In this example, the model contains two regressors:

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + U_i, \\
\mathrm{E}\left[X_{1,i} U_i\right] &= \mathrm{E}\left[X_{2,i} U_i\right] = 0.
\end{aligned}
$$

- However, since X_2 is not controlled for, it goes into the error term:

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_{1,i} + V_i, \text{ where} \\
V_i &= \beta_2 X_{2,i} + U_i.
\end{aligned}
$$

For consistency of $\tilde{\beta}_{1,n}$ we need $\mathrm{Cov}\left[X_{1,i}, V_i\right]$ to be equal to zero, however,

$$
\begin{aligned}
\mathrm{Cov}\left[X_{1,i}, V_i\right] &= \mathrm{Cov}\left[X_{1,i}, \beta_2 X_{2,i} + U_i\right] \\
&= \mathrm{Cov}\left[X_{1,i}, \beta_2 X_{2,i}\right] + \mathrm{Cov}\left[X_{1,i}, U_i\right] \\
&= \beta_2 \mathrm{Cov}\left[X_{1,i}, X_{2,i}\right] + 0 \\
&\neq 0, \text{ unless } \beta_2 = 0 \text{ or } \mathrm{Cov}\left[X_{1,i}, X_{2,i}\right] = 0.
\end{aligned}
$$