

# Lecture 17: Asymptotics

Economics 326 — Introduction to Econometrics II

Vadim Marmer, UBC

## Part I: Consistency

### Why we need large-sample theory

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

### Limitations of finite-sample theory

- If the errors are only **weakly exogenous**:

$$E[X_i U_i] = 0,$$

the OLS estimator is in general biased.

- If the errors are **heteroskedastic**:

$$E[U_i^2 | X_i] = h(X_i),$$

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.
- Asymptotic (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  $\theta_n$  be a sequence of random variables indexed by the sample size  $n$ . We say that  $\theta_n$  **converges in probability** if

$$\lim_{n \rightarrow \infty} P(|\theta_n - \theta| \geq \varepsilon) = 0 \text{ for all } \varepsilon > 0.$$

- We denote this as  $\theta_n \rightarrow_p \theta$  or  $p \lim \theta_n = \theta$ .
- An example of convergence in probability is a Law of Large Numbers (LLN):

Let  $X_1, X_2, \dots, X_n$  be a random sample such that  $E[X_i] = \mu$  for all  $i = 1, \dots, n$ , and define  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Then, under certain conditions,

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

## LLN

- Let  $X_1, \dots, X_n$  be a sample of **independent identically distributed (iid)** random variables. Let  $E[X_i] = \mu$ . If  $\text{Var}(X_i) = \sigma^2 < \infty$ , then

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

- In fact when the data are **iid**, the LLN holds if

$$E[|X_i|] < \infty,$$

but we prove the result under a stronger assumption that  $\text{Var}(X_i) < \infty$ .

## Markov's inequality

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

$$P(|W| \geq \varepsilon) \leq \frac{E[|W|^r]}{\varepsilon^r}.$$

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

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

- The probability of observing an outlier (a large deviation of  $X$  from its mean  $\mu$ ) can be bounded by the variance.

## Proof of the LLN

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

$$\begin{aligned} P(|\bar{X}_n - \mu| \geq \varepsilon) &= P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) \\ &= P\left(\left|\frac{1}{n} \sum_{i=1}^n (X_i - \mu)\right| \geq \varepsilon\right) \\ &\leq \frac{E\left[\left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)\right)^2\right]}{\varepsilon^2} \\ &= \frac{1}{n^2 \varepsilon^2} \left( \sum_{i=1}^n E[(X_i - \mu)^2] + \sum_{i=1}^n \sum_{j \neq i} E[(X_i - \mu)(X_j - \mu)] \right) \\ &= \frac{1}{n^2 \varepsilon^2} \left( \sum_{i=1}^n \text{Var}(X_i) + \sum_{i=1}^n \sum_{j \neq i} \text{Cov}(X_i, X_j) \right) \\ &= \frac{n\sigma^2}{n^2 \varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \text{ as } n \rightarrow \infty \text{ for all } \varepsilon > 0. \end{aligned}$$

## Averaging and variance reduction

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

$$\begin{aligned} \mathbb{E}[X_i] &= \mu \text{ for all } i = 1, \dots, n, \\ \text{Var}(X_i) &= \sigma^2 \text{ for all } i = 1, \dots, n, \\ \text{Cov}(X_i, X_j) &= 0 \text{ for all } j \neq i. \end{aligned}$$

- The mean of the sample average:

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

### Variance of the sample average

- The variance of the sample average:

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

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

### Convergence in probability: properties

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

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

- If  $\theta_n \rightarrow_p \theta$ , then  $\theta_n^2 \rightarrow_p \theta^2$ .
- If  $\theta_n \rightarrow_p \theta$  and  $\theta \neq 0$ , then  $1/\theta_n \rightarrow_p 1/\theta$ .

- Suppose that  $\theta_n \rightarrow_p \theta$  and  $\lambda_n \rightarrow_p \lambda$ . Then,

- $\theta_n + \lambda_n \rightarrow_p \theta + \lambda$ .
- $\theta_n \lambda_n \rightarrow_p \theta \lambda$ .
- $\theta_n / \lambda_n \rightarrow_p \theta / \lambda$  provided that  $\lambda \neq 0$ .

## Consistency

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

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

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

## Consistency of OLS

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

$$\hat{\beta}_{0,n} \rightarrow_p \beta_0,$$

$$\hat{\beta}_{1,n} \rightarrow_p \beta_1.$$

- The key identifying assumption is Assumption 3:  $\text{Cov}(X_i, U_i) = 0$ .

## Proof of consistency

- Write

$$\begin{aligned} \hat{\beta}_{1,n} &= \frac{\sum_{i=1}^n (X_i - \bar{X}_n) Y_i}{\sum_{i=1}^n (X_i - \bar{X}_n)^2} = \beta_1 + \frac{\sum_{i=1}^n (X_i - \bar{X}_n) U_i}{\sum_{i=1}^n (X_i - \bar{X}_n)^2} \\ &= \beta_1 + \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n) U_i}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}. \end{aligned}$$

- We will show that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n) U_i &\rightarrow_p 0, \\ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 &\rightarrow_p \text{Var}(X_i), \end{aligned}$$

- Since  $\text{Var}(X_i) \neq 0$ ,

$$\hat{\beta}_{1,n} = \beta_1 + \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n) U_i}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2} \rightarrow_p \beta_1 + \frac{0}{\text{Var}(X_i)} = \beta_1.$$

### Numerator converges to zero

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

By the LLN,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_i U_i &\rightarrow_p \mathbb{E}[X_i U_i] = 0, \\ \bar{X}_n &\rightarrow_p \mathbb{E}[X_i], \\ \frac{1}{n} \sum_{i=1}^n U_i &\rightarrow_p \mathbb{E}[U_i] = 0. \end{aligned}$$

Hence,

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

### Denominator converges to $\text{Var}(X_i)$

- First,

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

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

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \rightarrow_p \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = \text{Var}(X_i).$$

### Multiple regression

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

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

where  $\mathbb{E}[U_i] = 0$ .

- The key assumption is that the errors and regressors are uncorrelated:

$$\mathbb{E}[X_{1,i} U_i] = \dots = \mathbb{E}[X_{k,i} U_i] = 0.$$

## Omitted variables and OLS inconsistency

- Suppose that the true model has two regressors:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + U_i,$$

$$E[X_{1,i} U_i] = E[X_{2,i} U_i] = 0.$$

- Suppose that the econometrician includes **only**  $X_1$  in the regression when estimating  $\beta_1$ :

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

### Inconsistency proof: noise term

- From the previous slide:

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

- As before,

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

### Inconsistency proof: bias term

- From the previous slide:

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

- However,

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

## Omitted variable bias formula

- We have

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

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

## OVB through the composite error

- In this example, the model contains two regressors:

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

- 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  $\text{Cov}(X_{1,i}, V_i) = 0$ ; however,

$$\begin{aligned}\text{Cov}(X_{1,i}, V_i) &= \text{Cov}(X_{1,i}, \beta_2 X_{2,i} + U_i) \\ &= \text{Cov}(X_{1,i}, \beta_2 X_{2,i}) + \text{Cov}(X_{1,i}, U_i) \\ &= \beta_2 \text{Cov}(X_{1,i}, X_{2,i}) + 0 \\ &\neq 0, \text{ unless } \beta_2 = 0 \text{ or } \text{Cov}(X_{1,i}, X_{2,i}) = 0.\end{aligned}$$

## Part II: Asymptotic Normality

### Why do we need asymptotic normality?

- In the previous lectures, we showed that the OLS estimator has an **exact** normal distribution when the **errors are normally distributed**.
  - The same assumption is needed to show that the  $T$  statistic has a  $t$ -distribution and the  $F$  statistic has an  $F$ -distribution.
- In this lecture, we 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  $\chi^2$  distribution (approximately).

## Asymptotic normality

- Let  $W_n$  be a sequence of random variables indexed by the sample size  $n$ .
  - Typically,  $W_n$  will be a function of some estimator, such as  $W_n = \sqrt{n}(\hat{\beta}_n - \beta)$ .
- We say that  $W_n$  has an **asymptotically** normal distribution if its **CDF** converges to a **normal CDF**.
- Let  $W$  be any random variable with a normal  $N(0, \sigma^2)$  distribution. We say that  $W_n$  has an asymptotically normal distribution if for all  $x \in \mathbb{R}$ :

$$F_n(x) = P(W_n \leq x) \rightarrow P(W \leq x) = F(x) \text{ as } n \rightarrow \infty.$$

- We denote this as  $W_n \rightarrow_d W$  or  $W_n \rightarrow_d N(0, \sigma^2)$ .

## Convergence in distribution

- Asymptotic normality is an example of convergence in distribution.
- We say that a sequence of random variables  $W_n$  converges in distribution to  $W$  (denoted as  $W_n \rightarrow_d W$ ) if the CDF of  $W_n$  converges to the CDF of  $W$  at all points where the CDF of  $W$  is continuous.
- Convergence in distribution is convergence of the **CDFs**.

## Central Limit Theorem (CLT)

- An example of convergence in distribution is a CLT.
- Let  $X_1, \dots, X_n$  be a sample of **iid** random variables such that  $E[X_i] = 0$  and  $\text{Var}(X_i) = \sigma^2 > 0$  (finite). Then, as  $n \rightarrow \infty$ ,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow_d N(0, \sigma^2).$$

## CLT with non-zero mean

- For the CLT we impose 3 assumptions: **(1)** iid; **(2)** Mean zero; **(3)** Finite variance different from zero.
- If  $X_1, \dots, X_n$  are iid but  $E[X_i] = \mu \neq 0$ , then consider  $X_i - \mu$ . Since  $E[X_i - \mu] = 0$ , we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \rightarrow_d N(0, \text{Var}(X_i)).$$

Then

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

## CLT for the sample average

- From the previous slide:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) = \sqrt{n} (\bar{X}_n - \mu).$$

- Thus, the CLT can be stated as

$$\sqrt{n} (\bar{X}_n - \mu) \rightarrow_d N(0, \text{Var}(X_i)).$$

- By the LLN,

$$\bar{X}_n - \mu \rightarrow_p 0,$$

and

$$\text{Var}(\sqrt{n} (\bar{X}_n - \mu)) = n \text{Var}(\bar{X}_n) = n \frac{\text{Var}(X_i)}{n} = \text{Var}(X_i).$$

## Properties

- Suppose that  $W_n \rightarrow_d N(0, \sigma^2)$  and  $\theta_n \rightarrow_p \theta$ . Then,

$$\theta_n W_n \rightarrow_d \theta N(0, \sigma^2) \stackrel{d}{=} N(0, \theta^2 \sigma^2),$$

and

$$\theta_n + W_n \rightarrow_d \theta + N(0, \sigma^2) \stackrel{d}{=} N(\theta, \sigma^2).$$

- Suppose that  $Z_n \rightarrow_d Z \sim N(0, 1)$ . Then,

$$Z_n^2 \rightarrow_d Z^2 \equiv \chi_1^2.$$

- If  $W_n \rightarrow_d c = \text{constant}$ , then  $W_n \rightarrow_p c$ .

## Asymptotic normality of OLS

- Suppose that:

1. The data  $\{(Y_i, X_i) : i = 1, \dots, n\}$  are iid.
2.  $Y_i = \beta_0 + \beta_1 X_i + U_i$ , where  $\text{E}[U_i] = 0$ .
3.  $\text{E}[X_i U_i] = 0$ .
4.  $0 < \text{Var}(X_i) < \infty$ .
5.  $0 < \text{E}[(X_i - \text{E}[X_i])^2 U_i^2] < \infty$  and  $0 < \text{E}[U_i^2] < \infty$ .

- Let  $\hat{\beta}_{1,n}$  be the OLS estimator of  $\beta_1$ . Then,

$$\sqrt{n} (\hat{\beta}_{1,n} - \beta_1) \rightarrow_d N\left(0, \frac{\text{E}[(X_i - \text{E}[X_i])^2 U_i^2]}{(\text{Var}(X_i))^2}\right).$$

- $V = \frac{\mathbb{E}[(X_i - \mathbb{E}[X_i])^2 U_i^2]}{(\text{Var}(X_i))^2}$  is called the **asymptotic variance** of  $\hat{\beta}_{1,n}$ .

### Large-sample approximation for OLS

- Let  $\stackrel{a}{\sim}$  denote “approximately in large samples.”
- The asymptotic normality

$$\sqrt{n}(\hat{\beta}_{1,n} - \beta_1) \rightarrow_d N(0, V)$$

can be viewed as the following large-sample **approximation**:

$$\sqrt{n}(\hat{\beta}_{1,n} - \beta_1) \stackrel{a}{\sim} N(0, V),$$

or

$$\hat{\beta}_{1,n} \stackrel{a}{\sim} N(\beta_1, V/n).$$

### Proof: decomposition

Write

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

Now

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

and

$$\sqrt{n}(\hat{\beta}_{1,n} - \beta_1) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \bar{X}_n) U_i}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}.$$

### Proof: combining the limits

$$\sqrt{n}(\hat{\beta}_{1,n} - \beta_1) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \bar{X}_n) U_i}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}.$$

In Part I, we established

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \rightarrow_p \text{Var}(X_i).$$

We will show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \bar{X}_n) U_i \rightarrow_d N(0, \mathbb{E}[(X_i - \mathbb{E}[X_i])^2 U_i^2]),$$

so that

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{1,n} - \beta_1) &= \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \bar{X}_n) U_i}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2} \\ &\rightarrow_d \frac{N(0, \mathbb{E}[(X_i - \mathbb{E}[X_i])^2 U_i^2])}{\text{Var}(X_i)} \\ &\stackrel{d}{=} N\left(0, \frac{\mathbb{E}[(X_i - \mathbb{E}[X_i])^2 U_i^2]}{(\text{Var}(X_i))^2}\right). \end{aligned}$$

**Proof: numerator CLT**

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \bar{X}_n) U_i &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mathbb{E}[X_i] + \mathbb{E}[X_i] - \bar{X}_n) U_i \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) U_i + (\mathbb{E}[X_i] - \bar{X}_n) \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i. \end{aligned}$$

We have

$$\mathbb{E}[(X_i - \mathbb{E}[X_i]) U_i] = \mathbb{E}[X_i U_i] - \mathbb{E}[X_i] \mathbb{E}[U_i] = 0,$$

and  $0 < \mathbb{E}[(X_i - \mathbb{E}[X_i])^2 U_i^2] < \infty$ , so by the CLT,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) U_i \rightarrow_d N(0, \mathbb{E}[(X_i - \mathbb{E}[X_i])^2 U_i^2]).$$

**Proof: second term vanishes**

It is left to show that

$$(\mathbb{E}[X_i] - \bar{X}_n) \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i \rightarrow_p 0.$$

We have  $\mathbb{E}[U_i] = 0$  and  $0 < \mathbb{E}[U_i^2] < \infty$ . By the CLT,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n U_i \rightarrow_d N(0, \mathbb{E}[U_i^2]).$$

By the LLN,

$$\mathbb{E}[X_i] - \bar{X}_n \rightarrow_p 0.$$

Hence, the result follows.

## Part III: Asymptotic Variance

### Asymptotic variance

- In Part II, we showed that when the data are iid and the regressors are exogenous,

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + U_i, \\ E[U_i] &= E[X_i U_i] = 0, \end{aligned}$$

the OLS estimator of  $\beta_1$  is asymptotically normal:

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{1,n} - \beta_1) &\rightarrow_d N(0, V), \\ V &= \frac{E[(X_i - E[X_i])^2 U_i^2]}{(\text{Var}(X_i))^2}. \end{aligned}$$

- For hypothesis testing, we need a **consistent** estimator of the asymptotic variance  $V$ :

$$\hat{V}_n \rightarrow_p V.$$

### Simplifying $V$ under homoskedasticity

- Assume that the errors are **homoskedastic**:

$$E[U_i^2 | X_i] = \sigma^2 \text{ for all } X_i\text{'s.}$$

- In this case, the asymptotic variance can be simplified using the Law of Iterated Expectation:

$$\begin{aligned} E[(X_i - E[X_i])^2 U_i^2] &= E[E[(X_i - E[X_i])^2 U_i^2 | X_i]] \\ &= E[(X_i - E[X_i])^2 E[U_i^2 | X_i]] \\ &= E[(X_i - E[X_i])^2 \sigma^2] \\ &= \sigma^2 E[(X_i - E[X_i])^2] = \sigma^2 \text{Var}(X_i). \end{aligned}$$

### Estimating $V$ : method of moments

- Thus, when the errors are homoskedastic with  $E[U_i^2] = \sigma^2$ ,

$$V = \frac{E[(X_i - E[X_i])^2 U_i^2]}{(\text{Var}(X_i))^2} = \frac{\sigma^2 \text{Var}(X_i)}{(\text{Var}(X_i))^2} = \frac{\sigma^2}{\text{Var}(X_i)}.$$

- Let  $\hat{U}_i = Y_i - \hat{\beta}_{0,n} - \hat{\beta}_{1,n} X_i$ , where  $\hat{\beta}_{0,n}$  and  $\hat{\beta}_{1,n}$  are the OLS estimators of  $\beta_0$  and  $\beta_1$ .
- A consistent estimator for the asymptotic variance can be constructed using the **Method of Moments**:

$$\begin{aligned} \hat{\sigma}_n^2 &= \frac{1}{n} \sum_{i=1}^n \hat{U}_i^2, \\ \widehat{\text{Var}}(X_i) &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \text{ and} \\ \hat{V}_n &= \frac{\hat{\sigma}_n^2}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}. \end{aligned}$$

## Why the LLN does not apply directly

- From the previous slide:

$$\hat{V}_n = \frac{\hat{\sigma}_n^2}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}, \quad \hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n \hat{U}_i^2, \quad \hat{U}_i = Y_i - \hat{\beta}_{0,n} - \hat{\beta}_{1,n} X_i.$$

- When proving the consistency of OLS (Part I), we showed that

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \rightarrow_p \text{Var}(X_i),$$

and to establish  $\hat{V}_n \rightarrow_p V$ , we need to show that  $\hat{\sigma}_n^2 \rightarrow_p \sigma^2$ .

- The LLN **cannot** be applied directly to

$$\frac{1}{n} \sum_{i=1}^n \hat{U}_i^2$$

because the  $\hat{U}_i$ 's are not iid: they are **dependent** through  $\hat{\beta}_{0,n}$  and  $\hat{\beta}_{1,n}$ .

## Proof of $\hat{\sigma}_n^2 \rightarrow_p \sigma^2$

- First, write

$$\begin{aligned} \hat{U}_i &= Y_i - \hat{\beta}_{0,n} - \hat{\beta}_{1,n} X_i \\ &= (\beta_0 + \beta_1 X_i + U_i) - \hat{\beta}_{0,n} - \hat{\beta}_{1,n} X_i \\ &= U_i - (\hat{\beta}_{0,n} - \beta_0) - (\hat{\beta}_{1,n} - \beta_1) X_i. \end{aligned}$$

- Now,

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n \hat{U}_i^2 = \frac{1}{n} \sum_{i=1}^n (U_i - (\hat{\beta}_{0,n} - \beta_0) - (\hat{\beta}_{1,n} - \beta_1) X_i)^2.$$

## Completing the consistency proof

- We have

$$\begin{aligned} \hat{\sigma}_n^2 &= \frac{1}{n} \sum_{i=1}^n (U_i - (\hat{\beta}_{0,n} - \beta_0) - (\hat{\beta}_{1,n} - \beta_1) X_i)^2 \\ &= \frac{1}{n} \sum_{i=1}^n U_i^2 + (\hat{\beta}_{0,n} - \beta_0)^2 + (\hat{\beta}_{1,n} - \beta_1)^2 \frac{1}{n} \sum_{i=1}^n X_i^2 \\ &\quad - 2(\hat{\beta}_{0,n} - \beta_0) \frac{1}{n} \sum_{i=1}^n U_i - 2(\hat{\beta}_{1,n} - \beta_1) \frac{1}{n} \sum_{i=1}^n U_i X_i \\ &\quad + 2(\hat{\beta}_{0,n} - \beta_0)(\hat{\beta}_{1,n} - \beta_1) \frac{1}{n} \sum_{i=1}^n X_i. \end{aligned}$$

- By the LLN,

$$\frac{1}{n} \sum_{i=1}^n U_i^2 \rightarrow_p \text{E}[U_i^2] = \sigma^2.$$

- Because  $\hat{\beta}_{0,n}$  and  $\hat{\beta}_{1,n}$  are consistent,

$$\hat{\beta}_{0,n} - \beta_0 \rightarrow_p 0 \text{ and } \hat{\beta}_{1,n} - \beta_1 \rightarrow_p 0.$$

## Using $s^2$ instead of $\hat{\sigma}_n^2$

- Thus, when the errors are homoskedastic,

$$\hat{V}_n = \frac{\hat{\sigma}_n^2}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}, \text{ with } \hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n \hat{U}_i^2,$$

is a consistent estimator of  $V = \frac{\sigma^2}{\text{Var}(X_i)}$ .

- Similarly,

$$s^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{U}_i^2 \rightarrow_p \sigma^2,$$

and therefore

$$\hat{V}_n = \frac{s^2}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$$

is also a consistent estimator of  $V = \frac{\sigma^2}{\text{Var}(X_i)}$ .

- This version has an advantage over the one with  $\hat{\sigma}_n^2$ : in addition to being consistent,  $s^2$  is also an unbiased estimator of  $\sigma^2$  if the regressors are strongly exogenous.

## Asymptotic approximation

- The result  $\sqrt{n}(\hat{\beta}_{1,n} - \beta_1) \rightarrow_d N(0, V)$  is used as the following **approximation**:

$$\hat{\beta}_{1,n} \overset{a}{\sim} N\left(\beta_1, \frac{V}{n}\right),$$

where  $\overset{a}{\sim}$  denotes approximately in large samples. Thus, the variance of  $\hat{\beta}_{1,n}$  can be taken as approximately  $V/n$ .

- With  $\hat{V}_n = \frac{s^2}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$  we have

$$\frac{\hat{V}_n}{n} = \frac{s^2}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2} \cdot \frac{1}{n} = \frac{s^2}{\sum_{i=1}^n (X_i - \bar{X}_n)^2}.$$

## Connection to the exact result

- From the previous slide:

$$\frac{\hat{V}_n}{n} = \frac{s^2}{\sum_{i=1}^n (X_i - \bar{X}_n)^2}$$

- Thus, in the case of homoskedastic errors we have the following asymptotic approximation:

$$\hat{\beta}_{1,n} \overset{a}{\sim} N\left(\beta_1, \frac{s^2}{\sum_{i=1}^n (X_i - \bar{X}_n)^2}\right).$$

- In finite samples, we have the same result **exactly**, when the regressors are **strongly exogenous** and the errors are **normal**.

## Asymptotic $T$ -test

- Consider testing  $H_0 : \beta_1 = \beta_{1,0}$  vs  $H_1 : \beta_1 \neq \beta_{1,0}$ .
- Consider the behavior of the  $T$  statistic **under**  $H_0 : \beta_1 = \beta_{1,0}$ . Since

$$\sqrt{n}(\hat{\beta}_{1,n} - \beta_1) \rightarrow_d N(0, V) \text{ and } \hat{V}_n \rightarrow_p V,$$

we have

$$\begin{aligned} T &= \frac{\hat{\beta}_{1,n} - \beta_{1,0}}{\sqrt{\hat{V}_n/n}} = \frac{\sqrt{n}(\hat{\beta}_{1,n} - \beta_{1,0})}{\sqrt{\hat{V}_n}} \\ &\stackrel{H_0}{=} \frac{\sqrt{n}(\hat{\beta}_{1,n} - \beta_1)}{\sqrt{\hat{V}_n}} \\ &\rightarrow_d \frac{N(0, V)}{\sqrt{V}} \stackrel{d}{=} N(0, 1). \end{aligned}$$

## Asymptotic $T$ -test: rejection rule

- Under  $H_0 : \beta_1 = \beta_{1,0}$ ,

$$T = \frac{\hat{\beta}_{1,n} - \beta_{1,0}}{\sqrt{\hat{V}_n/n}} \rightarrow_d N(0, 1),$$

provided that  $\hat{V}_n \rightarrow_p V$  (the asymptotic variance of  $\hat{\beta}_{1,n}$ ).

- An asymptotic size  $\alpha$  test rejects  $H_0 : \beta_1 = \beta_{1,0}$  against  $H_1 : \beta_1 \neq \beta_{1,0}$  when

$$|T| > z_{1-\alpha/2},$$

where  $z_{1-\alpha/2}$  is a **standard normal** critical value.

- Asymptotically, the variance of the OLS estimator is known; we behave as if the variance were known.

## Heteroskedastic errors

- In general, the **errors are heteroskedastic**:  $E[U_i^2 | X_i]$  is not constant and changes with  $X_i$ .
- In this case,  $\hat{V}_n = \frac{s^2}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$  is **not** a consistent estimator of the asymptotic variance  $V = \frac{E[(X_i - E[X_i])^2 U_i^2]}{(\text{Var}(X_i))^2}$ :

$$\begin{aligned} \frac{s^2}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2} &\rightarrow_p \frac{E[U_i^2]}{\text{Var}(X_i)} \\ &= \frac{\text{Var}(X_i) \cdot E[U_i^2]}{(\text{Var}(X_i))^2} \\ &\neq \frac{E[(X_i - E[X_i])^2 U_i^2]}{(\text{Var}(X_i))^2}. \end{aligned}$$

## HC estimator of asymptotic variance

- In the case of heteroskedastic errors, a consistent estimator of  $V = \frac{E[(X_i - E[X_i])^2 U_i^2]}{(\text{Var}(X_i))^2}$  can be constructed as follows:

$$\hat{V}_n^{HC} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \hat{U}_i^2}{\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2\right)^2}.$$

- One can show that  $\hat{V}_n^{HC} \rightarrow_p V$  whether the errors are heteroskedastic or homoskedastic.
- We have the following asymptotic approximation:

$$\hat{\beta}_{1,n} \overset{a}{\sim} N\left(\beta_1, \frac{\hat{V}_n^{HC}}{n}\right),$$

and the standard errors can be computed as  $se(\hat{\beta}_{1,n}) = \sqrt{\hat{V}_n^{HC}/n}$ .

## HC standard errors in R

- In R, the HC estimator of standard errors can be obtained using the `sandwich` package:

```
library(wooldridge)
library(lmtest)
library(sandwich)
data("wage1")
reg <- lm(wage ~ educ + exper + tenure, data = wage1)
```

- **Standard (homoskedastic) standard errors:**

```
coeftest(reg)
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-2.872735	0.728964	-3.9408	9.225e-05	***
educ	0.598965	0.051284	11.6795	< 2.2e-16	***
exper	0.022340	0.012057	1.8528	0.06447	.
tenure	0.169269	0.021645	7.8204	2.935e-14	***
---					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

- **HC (robust) standard errors:**

```
coeftest(reg, vcov = vcovHC(reg, type = "HC1"))
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-2.872735	0.807415	-3.5579	0.0004078	***
educ	0.598965	0.061014	9.8169	< 2.2e-16	***
exper	0.022340	0.010555	2.1165	0.0347731	*
tenure	0.169269	0.029278	5.7814	1.277e-08	***
---					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					