Lecture 17: Asymptotics

Economics 326 — Introduction to Econometrics II

Vadim Marmer, UBC

April 5, 2026

Why we need large-sample theory

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

Limitations of finite-sample theory

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

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

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

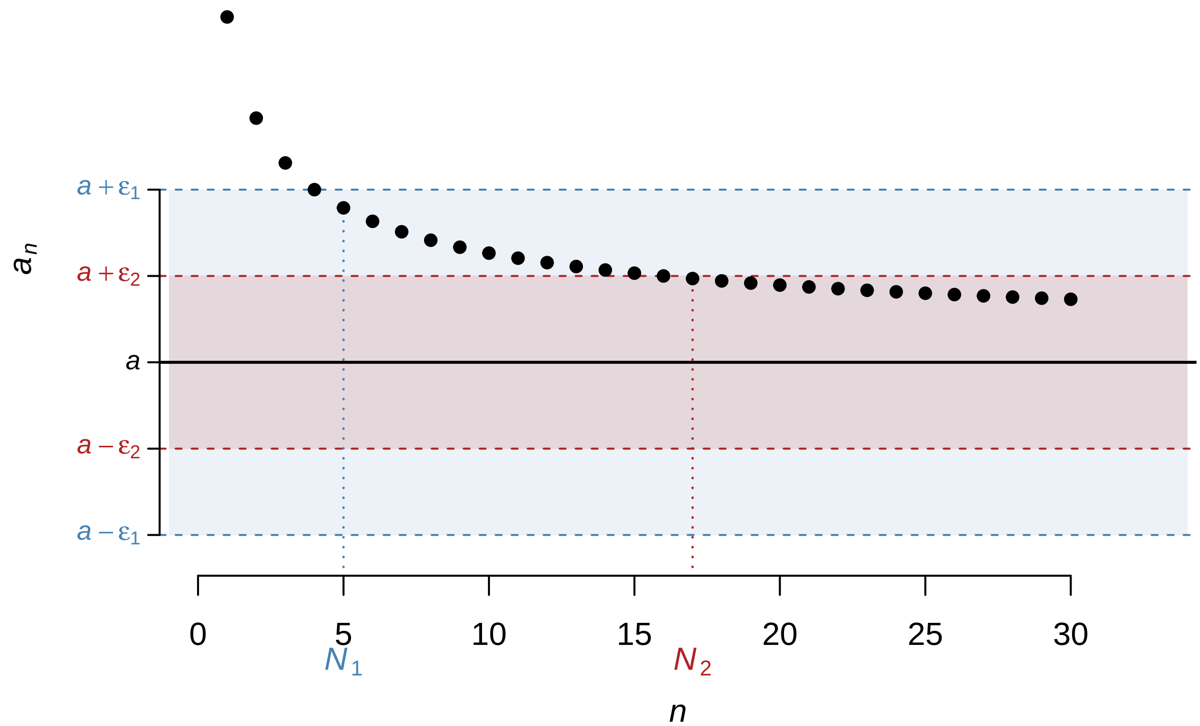
  $$\mathrm{E}[U_i^2 \mid X_i] = h(X_i),$$

  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 $\mathbf{X}$, 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.

Part I: Consistency

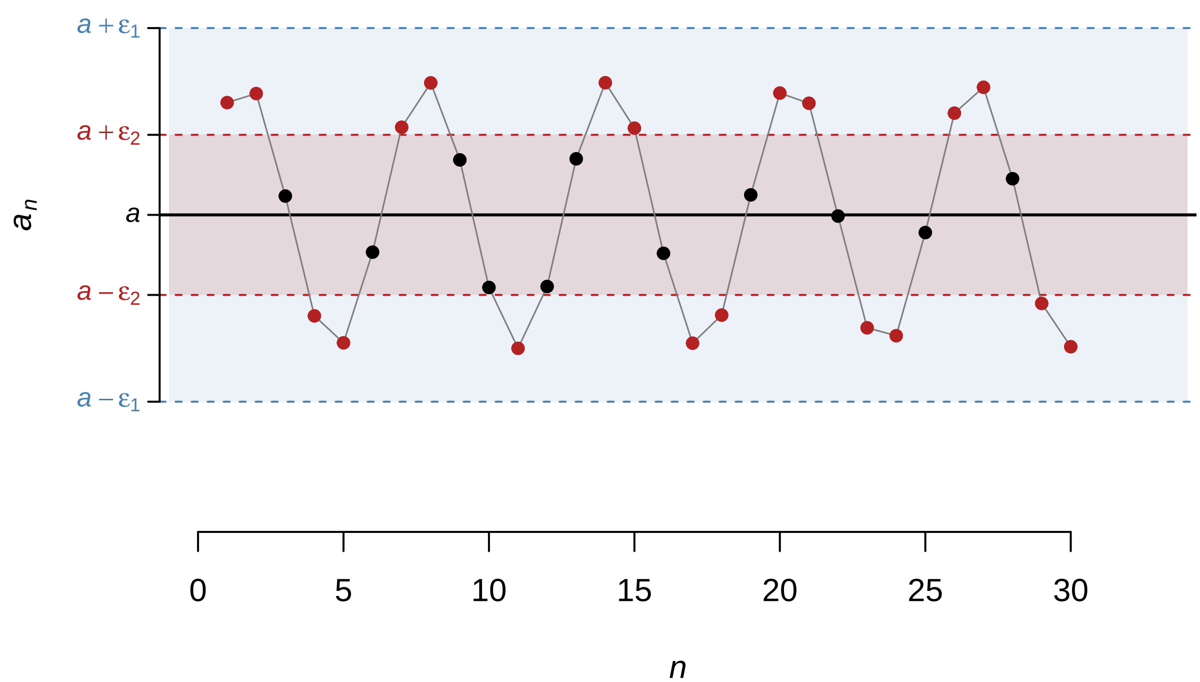
Convergence of a sequence

- A sequence of real numbers $a_1, a_2, ...$ **converges** to a if for every $\varepsilon > 0$ there exists N such that $|a_n - a| < \varepsilon$ for all $n \geq N$. We write $a_n \to a$.



Since $\varepsilon_1 > \varepsilon_2$, the ε_2-band is narrower, so it takes more terms for the sequence to stay inside it: $N_2 > N_1$. Smaller ε requires larger N.

- A sequence that **does not converge**: $a_n = a + c\sin(n)$ oscillates indefinitely around a.



For $\varepsilon_1 > c$, all terms lie within the ε_1-band. But for $\varepsilon_2 < c$, terms keep falling outside the ε_2-band (red dots) no matter how far along the sequence we go. Convergence requires the condition to hold for **all** $\varepsilon > 0$, so the sequence does not converge.

- Our estimator $\hat{\beta}_n$ is **random**: its value changes with each sample. To apply the concept of convergence, we need to convert it into a **non-random** sequence indexed by n.

- We take $a_n = P\left(\left|\hat{\beta}_n - \beta\right| \geq \varepsilon\right)$, which is a non-random number for each n. We say $\hat{\beta}_n$ **converges in probability** to β if $a_n \to 0$ for all $\varepsilon > 0$.

Convergence in probability and LLN

- More generally, let θ_n be a sequence of random variables indexed by the sample size n. We say that θ_n **converges in probability** to θ if

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

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

  Let $X_1, X_2, \ldots, X_n$ 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 $\mathrm{E}\left[X_i\right] = \mu$. If $\mathrm{Var}\left(X_i\right) = \sigma^2 < \infty$, then

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

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

$$\mathrm{E}\left[|X_i|\right] < \infty,$$

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

Markov's inequality

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

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

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

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

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

Proof of Markov's inequality

- For any event A, the expectation of its indicator equals the probability of the event:

$$\mathrm{E}\left[\mathbf{1}(A)\right] = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A).$$

- Define the indicator $\mathbf{1}\left(|W| \geq \varepsilon\right)$, which equals 1 when $|W| \geq \varepsilon$ and 0 otherwise. Then:

$$
\begin{aligned}
&\mathbf{1}\left(|W| \geq \varepsilon\right) \\
&\quad = \mathbf{1}\left(|W|^r \geq \varepsilon^r\right) \\
&\quad = \mathbf{1}\left(\frac{|W|^r}{\varepsilon^r} \geq 1\right) \\
&\quad \leq \frac{|W|^r}{\varepsilon^r} \\
&\implies \\
&P\left(|W| \geq \varepsilon\right) = \mathrm{E}\left[\mathbf{1}\left(|W| \geq \varepsilon\right)\right] \leq \frac{\mathrm{E}\left[|W|^r\right]}{\varepsilon^r}.
\end{aligned}
$$

Proof of the LLN

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

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

Averaging and variance reduction

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

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

- The mean of the sample average:

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

Variance of the sample average

- The variance of the sample average:

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

- 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,

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

  - 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 arbitrarily 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 $\mathrm{E}\left[U_i\right] = 0$.
  3. $\mathrm{E}\left[X_i U_i\right] = 0$.
  4. $0 < \operatorname{Var}\left(X_i\right) < \infty$.

- Let $\hat{\beta}_{0,n}$ and $\hat{\beta}_{1,n}$ be the OLS estimators of β_0 and β_1 based on a sample of size n. Under Assumptions 1–4,

$$\begin{aligned}\hat{\beta}_{0,n} &\to_p \beta_0,\\ \hat{\beta}_{1,n} &\to_p \beta_1.\end{aligned}$$

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

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

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

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

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

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,

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

Hence,

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

Denominator converges to $\mathrm{Var}(X_i)$

- The sample variance can be written as

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

- 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:

$$Y_i=\beta_0+\beta_1X_{1,i}+...+\beta_kX_{k,i}+U_i,$$

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

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

Omitted variables and OLS inconsistency

- Suppose that the true model has two regressors:

$$\begin{aligned}&Y_i=\beta_0+\beta_1X_{1,i}+\beta_2X_{2,i}+U_i,\\&\mathrm{E}\left[X_{1,i}U_i\right]=\mathrm{E}\left[X_{2,i}U_i\right]=0.\end{aligned}$$

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

$$\begin{aligned}\tilde{\beta}_{1,n}&=\frac{\sum_{i=1}^{n}\left(X_{1,i}-\bar{X}_{1,n}\right)Y_i}{\sum_{i=1}^{n}\left(X_{1,i}-\bar{X}_{1,n}\right)^2}\\&=\frac{\sum_{i=1}^{n}\left(X_{1,i}-\bar{X}_{1,n}\right)\left(\beta_0+\beta_1X_{1,i}+\beta_2X_{2,i}+U_i\right)}{\sum_{i=1}^{n}\left(X_{1,i}-\bar{X}_{1,n}\right)^2}\\&=\beta_1+\beta_2\frac{\sum_{i=1}^{n}\left(X_{1,i}-\bar{X}_{1,n}\right)X_{2,i}}{\sum_{i=1}^{n}\left(X_{1,i}-\bar{X}_{1,n}\right)^2}\\&\quad+\frac{\sum_{i=1}^{n}\left(X_{1,i}-\bar{X}_{1,n}\right)U_i}{\sum_{i=1}^{n}\left(X_{1,i}-\bar{X}_{1,n}\right)^2}.\end{aligned}$$

- Dividing numerator and denominator by n and applying the LLN as before:
  - The noise term vanishes:

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i}-\bar{X}_{1,n}\right)U_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i}-\bar{X}_{1,n}\right)^2}\to_p\frac{\mathrm{Cov}\left(X_{1,i},U_i\right)}{\mathrm{Var}\left(X_{1,i}\right)}=0.$$

- The bias term converges:

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i}-\bar{X}_{1,n}\right)X_{2,i}}{\frac{1}{n}\sum_{i=1}^{n}\left(X_{1,i}-\bar{X}_{1,n}\right)^{2}} \to_p \frac{\mathrm{Cov}\left(X_{1,i},X_{2,i}\right)}{\mathrm{Var}\left(X_{1,i}\right)}.$$

- Therefore,

$$\tilde{\beta}_{1,n} \to_p \beta_1 + \beta_2\frac{\mathrm{Cov}\left(X_{1,i},X_{2,i}\right)}{\mathrm{Var}\left(X_{1,i}\right)}.$$

- $\tilde{\beta}_{1,n}$ 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).

OVB through the composite error

- In this example, the model contains two regressors:

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

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

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

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 χ^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 $W_n = \sqrt{n}\left(\hat{\beta}_n - \beta\right)$.
- 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 $N\left(0, \sigma^2\right)$ distribution and let F denote its CDF. We say that W_n has an asymptotically normal distribution if for all $x \in \mathbb{R}$:

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

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

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 $W_n \to_d W$) 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 $X_1, \ldots, X_n$ be a sample of **iid** random variables such that $\mathrm{E}\left[X_i\right] = 0$ and $\mathrm{Var}\left(X_i\right) = \sigma^2 > 0$ (finite). Then, as $n \to \infty$,

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

- $\to_d$ means that the CDF of the scaled sum converges to the normal CDF: for every x,

$$P\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} X_i \leq x\right) \to \Phi(x) \text{ as } n \to \infty,$$

  where Φ is the standard normal CDF. For large n, the distribution of the scaled sum is approximately normal.

CLT with non-zero mean

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

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(X_i - \mu\right) \to_d N\left(0, \mathrm{Var}\left(X_i\right)\right).$$

  Then

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

CLT for the sample average

- From the previous slide:

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

- Thus, the CLT can be stated as

$$\sqrt{n}\left(\bar{X}_n-\mu\right)\to_d N\left(0,\mathrm{Var}\left(X_i\right)\right).$$

- By the LLN,

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

  and

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

Properties

- Suppose that $W_n\to_d N\left(0,\sigma^2\right)$ and $\theta_n\to_p\theta$. Then,

$$\theta_n W_n\to_d \theta N\left(0,\sigma^2\right)\overset{d}{=}N\left(0,\theta^2\sigma^2\right),$$

  and

$$\theta_n+W_n\to_d \theta+N\left(0,\sigma^2\right)\overset{d}{=}N\left(\theta,\sigma^2\right).$$

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

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

- If $W_n\to_d c=\text{constant}$, then $W_n\to_p c$.

Asymptotic normality of OLS

- Suppose that:
  1. The data $\{(Y_i,X_i):i=1,\ldots,n\}$ are iid.
  2. $Y_i=\beta_0+\beta_1X_i+U_i$, where $\mathrm{E}\left[U_i\right]=0$.
  3. $\mathrm{E}\left[X_iU_i\right]=0$.
  4. $0<\mathrm{Var}\left(X_i\right)<\infty$.
  5. $0<\mathrm{E}\left[\left(X_i-\mathrm{E}\left[X_i\right]\right)^2U_i^2\right]<\infty$ and $0<\mathrm{E}\left[U_i^2\right]<\infty$.

- Let $\hat{\beta}_{1,n}$ be the OLS estimator of β_1. Then,

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

- $V = \dfrac{\mathrm{E}\left[\left(X_i - \mathrm{E}\left[X_i\right]\right)^2 U_i^2\right]}{\left(\mathrm{Var}\left(X_i\right)\right)^2}$ 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

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

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

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

  or

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

Proof: decomposition

Write

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

Now

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

and

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

Proof: combining the limits

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

In Part I, we established

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

We will show that

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

so that

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

Proof: numerator CLT

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

We have

$$\mathrm{E}\left[\left(X_i-\mathrm{E}\left[X_i\right]\right)U_i\right]=\mathrm{E}\left[X_iU_i\right]-\mathrm{E}\left[X_i\right]\mathrm{E}\left[U_i\right]=0,$$

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

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

Proof: second term vanishes

It is left to show that

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

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

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i \to_d N\left(0, \mathrm{E}\left[U_i^2\right]\right).$$

By the LLN,

$$\mathrm{E}\left[X_i\right]-\bar{X}_n \to_p 0.$$

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,

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

  the OLS estimator of β_1 is asymptotically normal:

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

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

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

Simplifying V under homoskedasticity

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

$$
\mathrm{E}\left[U_i^2 \mid X_i\right] = \sigma^2 \text{ for all } X_i\text{'s}.
$$

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

$$
\begin{aligned}
\mathrm{E}\left[\left(X_i - \mathrm{E}\left[X_i\right]\right)^2 U_i^2\right] &= \mathrm{E}\left[\mathrm{E}\left[\left(X_i - \mathrm{E}\left[X_i\right]\right)^2 U_i^2 \mid X_i\right]\right] \\
&= \mathrm{E}\left[\left(X_i - \mathrm{E}\left[X_i\right]\right)^2 \mathrm{E}\left[U_i^2 \mid X_i\right]\right] \\
&= \mathrm{E}\left[\left(X_i - \mathrm{E}\left[X_i\right]\right)^2 \sigma^2\right] \\
&= \sigma^2\, \mathrm{E}\left[\left(X_i - \mathrm{E}\left[X_i\right]\right)^2\right] = \sigma^2 \mathrm{Var}\left(X_i\right).
\end{aligned}
$$

Estimating V: method of moments

- Thus, when the errors are homoskedastic with $\mathrm{E}\left[U_i^2\right] = \sigma^2$,

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

- 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 β_0 and β_1.
- A consistent estimator for the asymptotic variance can be constructed using the **Method of Moments**:

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

Why the LLN does not apply directly

- From the previous slide:

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

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

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

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

- 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: $\hat{\sigma}_n^2 \to_p \sigma^2$

- First, write

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

- Now,

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

Completing the consistency proof

- We have

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

- By the LLN,

$$\frac{1}{n}\sum_{i=1}^n U_i^2 \to_p \mathrm{E}\left[U_i^2\right] = \sigma^2.$$

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

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

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

- Thus, when the errors are homoskedastic,

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

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

- Similarly,

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

  and therefore

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

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

- 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 σ^2 if the regressors are strongly exogenous.

Asymptotic approximation

- The result $\sqrt{n}\left(\hat{\beta}_{1,n} - \beta_1\right) \to_d N(0, V)$ 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 $\hat{V}_n = \frac{s^2}{\frac{1}{n}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2}$ we have

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

Connection to the exact result

- From the previous slide:

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

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

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

- 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

$$\sqrt{n}\left(\hat{\beta}_{1,n} - \beta_1\right) \to_d N(0, V) \text{ and } \hat{V}_n \to_p V,$$

we have

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

Asymptotic T-test: rejection rule

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

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

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

- An asymptotic size α 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**: $\mathrm{E}\left[U_i^2 \mid X_i\right]$ 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 $V = \frac{\mathrm{E}\left[(X_i - \mathrm{E}[X_i])^2 U_i^2\right]}{(\mathrm{Var}(X_i))^2}$:

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

HC estimator of asymptotic variance

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

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

- One can show that $\hat{V}_n^{HC} \to_p V$ 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 $\text{se}\left(\hat{\beta}_{1,n}\right) = \sqrt{\hat{V}_n^{HC}/n}$.

HC standard errors in R

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

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

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

```r
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:**

```r
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
```