Econometrics II

OLS - Small and Large Sample Properties. Hypothesis Testing.

Lasha Chochua

2026

Finite Sample Properties

Introduction

- In the first part of the lecture, we investigate some **finite-sample** properties of the least squares estimator in the linear regression model.
- We calculate its **finite-sample** expectation and covariance matrix.
- Standard errors for the coefficient estimators are proposed.

Random Sampling

- To derive the **finite-sample** properties of the estimators, we need to specify the dependence structure across observations.
- The simplest case: observations are **independent and identically distributed (i.i.d.)**.
- Another term: **random sample**.

Random Sampling

Assumption 1: Random Sampling

The random variables

$$\{(Y_1, X_1), \dots, (Y_i, X_i), \dots, (Y_n, X_n)\}$$

are independent and identically distributed.

- Implications of **Assumption 1**:
  - If $i \neq j$, then (Y_i, X_i) is independent of (Y_j, X_j), but both follow the same distribution.
  - Independence means that decisions of individual i do not affect those of individual j, and vice versa.

When Does Assumption 1 Fail?

- If individuals are **connected**, independence may not hold.
  - Examples:
    - Neighbors
    - Members of the same village
    - Classmates
    - Firms in the same industry
- Interconnected decisions introduce mutual dependence.

Clustered Dependence

- A popular approach to account for dependence is **clustered dependence**.
- Observations are grouped into **clusters** (e.g., schools).

Definition of Unbiased Estimator

Definition 1: Unbiased Estimator

An estimator $\widehat{\theta}$ for θ is **unbiased** if

$$E[\widehat{\theta}] = \theta.$$

Assumption 2 – Linear Regression Model

Assumption 2: Linear Regression Model

The variables (Y, X) satisfy the linear regression equation:

$$Y = X'\beta + e \tag{1}$$

$$E[e \mid X] = 0 \tag{2}$$

The variables have finite second moments: $E[Y^2] < \infty$, $E[\|X\|^2] < \infty$, and an **invertible design matrix**: $\mathbf{Q}_{XX} = E[XX'] > 0$.

Note: For $X = (X_1, \dots, X_k)'$, the squared Euclidean norm is $\|X\|^2 = X'X = \sum_{j=1}^{k} X_j^2$, so $E[\|X\|^2] = \sum_{j=1}^{k} E[X_j^2] < \infty$ means each regressor has a finite second moment.

Heteroskedasticity vs. Homoskedasticity

- We will consider both:
  - The **general case** of heteroskedastic regression, where the conditional variance $E[e^2 \mid X] = \sigma^2(X)$ is **unrestricted**.
  - The **special case** of homoskedastic regression, where the conditional variance is **constant**.
- In the latter case, we add the following assumption:

Assumption 3: Homoskedastic Linear Regression Model

In addition to Assumption 2:

$$E[e^2 \mid X] = \sigma^2(X) = \sigma^2 \quad (3)$$

is independent of X.

Expectation of Least Squares Estimator

- Now we show that the **OLS estimator is unbiased** in the linear regression model.
- This calculation can be done using either:
  - **Summation notation**
  - **Matrix notation**
- We will use both approaches.

Conditional Expectation of Y_i

- Under (1)–(2), we observe:

$$E[Y_i \mid X_1, \ldots, X_n] = E[Y_i \mid X_i] = X_i'\beta \tag{4}$$

- The first equality states that the conditional expectation of Y_i given $\{X_1, \ldots, X_n\}$ only depends on X_i because the observations are **independent across** i.
- The second equality follows from the assumption of a **linear conditional expectation**.

Expectation of the OLS Estimator

$$E[\hat{\beta} \mid X_1, \dots, X_n] = E\left[\left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \sum_{i=1}^{n} X_i Y_i \mid X_1, \dots, X_n\right]$$

$$= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} E\left[\sum_{i=1}^{n} X_i Y_i \mid X_1, \dots, X_n\right]$$

$$= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \sum_{i=1}^{n} E\left[X_i Y_i \mid X_1, \dots, X_n\right]$$

Expectation of the OLS Estimator (Cont.)

$$
\begin{aligned}
&= \left(\sum_{i=1}^{n} X_i X_i' \right)^{-1} \sum_{i=1}^{n} X_i E[Y_i \mid X_1, \dots, X_n] \\
&= \left(\sum_{i=1}^{n} X_i X_i' \right)^{-1} \sum_{i=1}^{n} X_i E[Y_i \mid X_i] \\
&= \left(\sum_{i=1}^{n} X_i X_i' \right)^{-1} \sum_{i=1}^{n} X_i X_i' \beta \\
&= \beta.
\end{aligned}
$$

- Thus, we conclude that **the OLS estimator is unbiased**.

Expectation of Least Squares Estimator – Matrix Notation

- Now let's show the same result using **matrix notation**.
- From (4), we get:

$$E[\mathbf{Y} \mid \mathbf{X}] = \begin{bmatrix} E[Y_1 \mid X] \\ \vdots \\ E[Y_i \mid X] \\ \vdots \end{bmatrix} = \begin{bmatrix} X_1'\beta \\ \vdots \\ X_i'\beta \\ \vdots \end{bmatrix} = \mathbf{X}\beta \tag{5}$$

- Similarly:

$$E[\mathbf{e} \mid \mathbf{X}] = \begin{bmatrix} E[e_1 \mid X] \\ \vdots \\ E[e_i \mid X] \\ \vdots \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = 0.$$

Expectation of $\hat{\beta}$

- Using the OLS estimator: $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$
- Applying the **conditioning theorem**, **linearity of expectations**, and (5):

$$\begin{aligned} E[\hat{\beta} \mid \mathbf{X}] &= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \mid \mathbf{X}\right] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{Y} \mid \mathbf{X}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta = \beta. \end{aligned}$$

Alternative Derivation

- Inserting $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$ into the formula for $\hat{\beta}$:

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{e}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} \qquad (6)\end{aligned}$$

Expectation of OLS Decomposition

- Equation (6) provides a **linear decomposition** of $\hat{\beta}$ into:
  - The true parameter β.
  - The stochastic component $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}$.
- Taking expectations:

$$E[\hat{\beta} - \beta \mid \mathbf{X}] = E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} \mid \mathbf{X}\right]$$
$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{e} \mid \mathbf{X}] = 0.$$

Theorem 1 – Expectation of Least Squares Estimator

Important

Theorem 1: Expectation of Least Squares Estimator

In the linear regression model (Assumption 2) with i.i.d. sampling (Assumption 1):

$$E[\widehat{\beta} \mid \mathbf{X}] = \beta \tag{7}$$

Interpretation of (7)

- Equation (7) states that the estimator $\hat{\beta}$ is **unbiased** for β, conditional on $\mathbf{X}$.
- This means that:
  - The **conditional distribution** of $\hat{\beta}$ is **centered** at β.
  - "Conditional on $\mathbf{X}$" means that the distribution is unbiased **for any realization of the regressor matrix** $\mathbf{X}$.
  - In conditional models, we simply say **"$\hat{\beta}$ is unbiased for β."**

Implicit Assumptions

- It is worth noting that (7) and **all finite sample results** discussed assume that:

$$\mathbf{X}'\mathbf{X} \text{ is full rank.}$$

Variance of Least Squares Estimator

- Now, we calculate the **conditional variance** of the OLS estimator.
- For any $r \times 1$ random vector Z, define the $r \times r$ **covariance matrix**:

$$\text{var}[Z] = E\left[(Z - E[Z])(Z - E[Z])'\right] = E[ZZ'] - (E[Z])(E[Z])'$$

- For any pair (Z, X), define the **conditional covariance matrix**:

$$\text{var}[Z \mid X] = E\left[(Z - E[Z \mid X])(Z - E[Z \mid X])' \,\middle|\, X\right].$$

Definition of $\mathbf{V}_{\hat{\beta}}$

- We define:

$$\mathbf{V}_{\hat{\beta}} \stackrel{\text{def}}{=} \text{var}[\hat{\beta} \mid \mathbf{X}]$$

as the **conditional covariance matrix** of the **regression coefficient estimators**.

- We now derive its form.

Variance of Regression Error

- The **conditional covariance matrix** of the $n \times 1$ **regression error** $\mathbf{e}$ is the $n \times n$ matrix:

$$\operatorname{var}[\mathbf{e} \mid \mathbf{X}] = E[\mathbf{e}\mathbf{e}' \mid \mathbf{X}] \stackrel{\text{def}}{=} \mathbf{D}.$$

Structure of $\mathbf{D}$

- The i^{th} diagonal element of $\mathbf{D}$ is:

$$E[e_i^2 \mid \mathbf{X}] = E[e_i^2 \mid X_i] = \sigma_i^2.$$

- The $(i, j)^{th}$ **off-diagonal element** of $\mathbf{D}$ is:

$$E[e_i e_j \mid \mathbf{X}] = E(e_i \mid X_i) E(e_j \mid X_j) = 0.$$

- **why?**

Structure of $\mathbf{D}$

- The i^{th} diagonal element of $\mathbf{D}$ is:

$$E[e_i^2 \mid \mathbf{X}] = E[e_i^2 \mid X_i] = \sigma_i^2.$$

- The $(i,j)^{th}$ **off-diagonal element** of $\mathbf{D}$ is:

$$E[e_i e_j \mid \mathbf{X}] = E(e_i \mid X_i)E(e_j \mid X_j) = 0.$$

- **why?**
- The **first equality** follows from the **zero conditional mean assumption for errors** (Assumption 2, equation (2)).
- The **second equality** follows from **independence assumption**.

Representation of $\mathbf{D}$

- Thus, $\mathbf{D}$ is a **diagonal matrix** with i^{th} diagonal element σ_i^2:

$$\mathbf{D} = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2) = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} \tag{8}$$

Variance of Least Squares Estimator (Cont.)

- In the **special case** of the **homoskedastic regression model** (3), we have:

$$E[e_i^2 \mid X_i] = \sigma_i^2 = \sigma^2$$

- This simplifies the **error covariance matrix** to:

$$\mathbf{D} = \mathbf{I}_n \sigma^2.$$

- In general, however, $\mathbf{D}$ **need not take this simplified form**.

General Case for Any Matrix $\mathrm{A} = \mathrm{A}(\mathrm{X})$

- For any $n \times r$ **matrix** $\mathbf{A}$:

$$\mathrm{var}[\mathbf{A}'\mathbf{Y} \mid \mathbf{X}] = \mathrm{var}[\mathbf{A}'\mathbf{e} \mid \mathbf{X}] = \mathbf{A}'\mathbf{D}\mathbf{A} \tag{9}$$

- In particular, we can write:

$$\hat{\beta} = \mathbf{A}'\mathbf{Y}, \quad \text{where} \quad \mathbf{A} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

- Thus, applying (9):

$$\mathbf{V}_{\hat{\beta}} = \mathrm{var}[\hat{\beta} \mid \mathbf{X}] = \mathbf{A}'\mathbf{D}\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

Weighted Version of $\mathbf{X}^{'}\mathbf{X}$

- It is useful to note:

$$\mathbf{X}'\mathbf{D}\mathbf{X} = \sum_{i=1}^{n} X_i X_i' \sigma_i^2.$$

- This represents a **weighted version** of $\mathbf{X}'\mathbf{X}$.

Special Case – Homoskedastic Regression

- If the **homoskedastic regression model** (3) holds, then:

$$\mathbf{D} = \mathbf{I}_n \sigma^2.$$

- So:

$$\mathbf{X}'\mathbf{D}\mathbf{X} = \mathbf{X}'\mathbf{X}\sigma^2,$$

- The **covariance matrix** then simplifies to:

$$\mathbf{V}_{\hat{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2.$$

Theorem 2 – Variance of Least Squares Estimator

Important

Theorem 2: Variance of Least Squares Estimator
In the linear regression model (Assumption 2) with i.i.d. sampling (Assumption 1):

$$\mathbf{V}_{\hat{\beta}} = \operatorname{var}[\widehat{\beta} \mid \mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{D}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \tag{10}$$

where $\mathbf{D}$ is defined in (8). If in addition the **error is homoskedastic** (Assumption 3), then (10) simplifies to:

$$\mathbf{V}_{\hat{\beta}} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Gauss-Markov Theorem

- The **Gauss-Markov Theorem** is one of the most celebrated results in econometric theory.
- It provides a **classical justification** for the **least squares estimator**, showing that it has the **lowest variance among unbiased estimators**.

Homoskedastic Linear Regression Model

- Write the **homoskedastic linear regression model** in vector format as:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e} \tag{11}$$

$$E[\mathbf{e} \mid \mathbf{X}] = 0 \tag{12}$$

$$\mathrm{var}[\mathbf{e} \mid \mathbf{X}] = \mathbf{I}_n \sigma^2 \tag{13}$$

Theorem 3 – Gauss-Markov

Important

Theorem 3: Gauss-Markov

Take the **homoskedastic linear regression model** (11)–(13). If $\tilde{\beta}$ is any **unbiased estimator** of β , then:

$$\operatorname{var}[\tilde{\beta} \mid \mathbf{X}] \geq \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}.$$

Interpretation of Theorem 3

- **Provides a lower bound** on the covariance matrix of **unbiased estimators** under homoskedasticity.
- It states that **no unbiased estimator** can have a **variance matrix smaller** than:

$$\sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

- Since the **variance of the OLS estimator** is exactly equal to this bound, **no unbiased estimator** has a **lower variance** than OLS.
- **Consequently, we describe OLS as an efficient estimator in the class of unbiased estimators**.

Historical Background

- Theorem 3 was first **articulated by Carl Friedrich Gauss** in **1823**.
- **Andrei Markov** provided a **textbook treatment** in **1912**, clarifying the **role of unbiasedness**, which Gauss had assumed **implicitly**.

BLUE vs. BUE

- The **classical version** of the theorem led to the description of OLS as the **Best Linear Unbiased Estimator (BLUE)**.
- However, **Theorem 3 shows that OLS is not just BLUE but the Best Unbiased Estimator (BUE)**.

BLUE vs. BUE

- The **classical version** of the theorem led to the description of OLS as the **Best Linear Unbiased Estimator (BLUE)**.
- However, **Theorem 3 shows that OLS is not just BLUE but the Best Unbiased Estimator (BUE)**.
- **So far:** OLS is BUE under Equations 11-13 – no distributional assumption needed for point estimation
- **Next:** hypothesis testing and confidence intervals require knowing the **distribution** of $\hat{\beta}$, which in small samples depends on the distribution of $\mathbf{e}$

Normal Regression

Assumption 4 – Normality of Error

Assumption 4: Normality of Error

Conditional on $\mathbf{X}$, the error vector $\mathbf{e}$ is normally distributed:

$$\mathbf{e} \mid \mathbf{X} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n).$$

- To derive the **exact finite-sample distribution** of $\widehat{\beta}$, we need a distributional assumption on the error $\mathbf{e}$ – not just its mean and variance.
- In combination with earlier assumptions, this implies:

$$\widehat{\beta} \mid \mathbf{X} \sim \mathcal{N}(\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}).$$

Assumption 4 – Joint Normality and the iid Structure

- The covariance matrix $\sigma^2 \mathbf{I}_n$ is **diagonal** $\Rightarrow$ errors are uncorrelated; joint normality then implies **independence**
- Errors are also **identically distributed**: each $e_i \sim \mathcal{N}(0, \sigma^2)$
- Together: **iid** $\Rightarrow$ joint density factorizes into a product of identical marginals:

$$f(\mathbf{e} \mid \mathbf{X}) = \prod_{i=1}^{n} f(e_i \mid \mathbf{X}) = \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{e_i^2}{2\sigma^2}} \right]^n$$

- This is the **iid principle**: joint $= [F(\cdot)]^n$ – the $\mathbf{I}_n$ structure is exactly what makes it factorize
- **Contrast:** if $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \Omega)$ with $\Omega \neq \mathbf{I}_n$, off-diagonal elements prevent factorization $\Rightarrow$ GLS

Theorem 4 – Linear Transformation of Normal Vector

Important

Theorem 4: Linear Transformation of Normal Vector
If $X \sim \mathcal{N}(\mu, \Sigma)$ and $Y = a + BX$, then

$$Y \sim \mathcal{N}(a + B\mu, B\Sigma B').$$

- This result helps derive the distribution of linear estimators like $\widehat{\beta}$.
- It shows how **affine transformations preserve normality**.

Finite-Sample Distribution of $\hat{\beta}$

- Under Assumptions 1–4, from (6):

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}$$

- Since $\mathbf{e} \sim \mathcal{N}(0, \sigma^2 I_n)$, applying Theorem 4:

$$\hat{\beta} \mid \mathbf{X} \sim \mathcal{N}\left(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$$

- Hence, $\hat{\beta}$ is **normally distributed** with known variance up to σ^2.

Residuals and Estimator of σ^2

- Define the residuals:

$$\hat{e}_i = Y_i - X_i'\hat{\beta}$$

- The residual sum of squares (RSS) is:

$$\text{RSS} = \sum_{i=1}^{n} \hat{e}_i^2 = \hat{\mathbf{e}}'\hat{\mathbf{e}}$$

- Unbiased estimator of σ^2:

$$\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{n} \hat{e}_i^2$$

Standard Error of $\hat{\beta}_j$

- From (10), the sampling variance under homoskedasticity is:

$$\text{var}[\hat{\beta}_j \mid \mathbf{X}] = \sigma^2 \cdot [(\mathbf{X}'\mathbf{X})^{-1}]_{jj}$$

- We estimate it using $\hat{\sigma}^2$:

$$\widehat{\text{se}}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 \cdot [(\mathbf{X}'\mathbf{X})^{-1}]_{jj}}$$

- Used for inference on individual coefficients.

Hypothesis Testing under the Normal Regression

Definition 2 – t-Statistic

Definition 2: t-Statistic

The t-statistic for testing $H_0 : \beta_j = \beta_{j,0}$ is:

$$t_j = \frac{\hat{\beta}_j - \beta_{j,0}}{\widehat{\text{se}}(\hat{\beta}_j)} \tag{14}$$

- Under H_0 and Assumptions 1–4:

$$t_j \sim t_{n-k}$$

- Used to test hypotheses about individual coefficients.

Definition 3 – t-Distribution

Definition 3: t-Distribution

Let $Z \sim \mathcal{N}(0, 1)$ and $U \sim \chi^2_\nu$ independently. Then the random variable

$$T = \frac{Z}{\sqrt{U/\nu}}$$

has a t-distribution with ν degrees of freedom: $T \sim t_\nu$.

t-Test Decision Rule

- For a two-sided test with significance level α:
  - Reject H_0 if:

$$|t_j| > t_{n-k,1-\alpha/2}$$

- For a one-sided test:
  - Reject H_0 if $t_j > t_{n-k,1-\alpha}$ (right tail)
  - Reject H_0 if $t_j < -t_{n-k,1-\alpha}$ (left tail)
- $t_{n-k,q}$: quantile of t-distribution with $n-k$ degrees of freedom.

Definition 4 – F-Distribution

Definition 4: F-Distribution

Let $U_1 \sim \chi^2_{d_1}$ and $U_2 \sim \chi^2_{d_2}$ independently. Then the random variable

$$F = \frac{(U_1/d_1)}{(U_2/d_2)}$$

has an F-distribution with (d_1, d_2) degrees of freedom: $F \sim F_{d_1,d_2}$.

F-Test – Joint Hypothesis (Homoskedasticity-only F-statistic)

- Consider testing:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_q = 0 \quad \text{vs.} \quad H_1 : \text{at least one } \beta_j \neq 0$$

- The F-statistic is:

$$F = \frac{(\text{RSS}_r - \text{RSS}_u)/q}{\text{RSS}_u/(n-k-1)}$$

where RSS_r is RSS from the restricted model, RSS_u from the unrestricted model, q is the number of restrictions, k the number of regressors, and n the sample size.

- Under H_0:

$$F \sim F_{k,n-k-1}$$

Large Sample Properties

Assumption 1 (Restated)

Assumption 1: Regularity Conditions

1. The variables (Y_i, X_i), $i = 1, \ldots, n$, are i.i.d.
2. $E[Y^2] < \infty$.
3. $E[\|X\|^2] < \infty$.
4. $\mathbf{Q}_{XX} = E[XX']$ is positive definite.

- $E[\|X\|^2] = E[X'X] = \sum_{j=1}^{k} E[X_j^2]$, i.e., all regressors have finite second moments.

Linear Regression Model – Large Sample

Assumption 5 (Large Sample): Linear Regression Model

The variables (Y, X) satisfy:

$$Y = X'\beta + e, \qquad E[Xe] = 0.$$

The variables have finite second moments: $E[Y^2] < \infty$, $E[\|X\|^2] < \infty$, and an **invertible design matrix**: $Q_{XX} = E[XX'] > 0$.

Consistency of Least Squares Estimator

- We use:
  - **Weak Law of Large Numbers (WLLN)**.
  - **Continuous Mapping Theorem (CMT)**.
- Goal: show that $\hat{\beta}$ is **consistent**.
- Three key steps:
  - OLS is a function of sample moments.
  - **WLLN**: sample moments converge in probability to population moments.
  - **CMT**: continuous functions preserve convergence in probability.

The OLS Estimator

- The OLS estimator is given by:

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} X_i Y_i \right) = \hat{\mathbf{Q}}_{XX}^{-1} \hat{\mathbf{Q}}_{XY}$$

- The terms:
  - $\hat{\mathbf{Q}}_{XX} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i'$
  - $\hat{\mathbf{Q}}_{XY} = \frac{1}{n} \sum_{i=1}^{n} X_i Y_i$

Weak Law of Large Numbers (WLLN)

- Under Assumption 1, the **WLLN** implies:

$$\hat{\mathbf{Q}}_{XX} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \xrightarrow{p} E[XX'] = \mathbf{Q}_{XX} \tag{15}$$

$$\hat{\mathbf{Q}}_{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \xrightarrow{p} E[XY] = \mathbf{Q}_{XY}.$$

- Sample moments **converge in probability** to population moments.

Consistency of $\hat{\beta}$

- By the **Continuous Mapping Theorem (CMT)**:
  - If $\hat{\mathbf{Q}}_{XX} \xrightarrow{p} \mathbf{Q}_{XX}$ and $\hat{\mathbf{Q}}_{XY} \xrightarrow{p} \mathbf{Q}_{XY}$, then:

$$\hat{\beta} = g(\hat{\mathbf{Q}}_{XX}, \hat{\mathbf{Q}}_{XY}) \xrightarrow{p} g(\mathbf{Q}_{XX}, \mathbf{Q}_{XY}) = \mathbf{Q}_{XX}^{-1}\mathbf{Q}_{XY} = \beta \tag{16}$$

- The **CMT ensures** that OLS inherits convergence in probability.

Application of the CMT

- We express $\hat{\beta} = g(\hat{\mathbf{Q}}_{XX}, \hat{\mathbf{Q}}_{XY})$ where $g(\mathbf{A}, \mathbf{b}) = \mathbf{A}^{-1}\mathbf{b}$.
- The function $g(\mathbf{A}, \mathbf{b})$ **is continuous** whenever $\mathbf{A}^{-1}$ exists.
- **Assumption 1** states that $\mathbf{Q}_{XX}$ is **positive definite**, ensuring $\mathbf{Q}_{XX}^{-1}$ exists.
- This justifies the application of the CMT in (16).

Alternative Demonstration

- Recall that:

$$\hat{\beta} - \beta = \hat{\mathbf{Q}}_{XX}^{-1} \hat{\mathbf{Q}}_{Xe} \tag{17}$$

  where $\hat{\mathbf{Q}}_{Xe} = \frac{1}{n} \sum_{i=1}^{n} X_i e_i$.

- By the **WLLN** and Assumption 5:

$$\hat{\mathbf{Q}}_{Xe} \xrightarrow{p} E[Xe] = 0.$$

- Therefore, from (17):

$$\hat{\beta} - \beta = \hat{\mathbf{Q}}_{XX}^{-1} \hat{\mathbf{Q}}_{Xe} \xrightarrow{p} \mathbf{Q}_{XX}^{-1} \cdot 0 = 0,$$

  i.e., $\hat{\beta} \xrightarrow{p} \beta$.

Theorem 5 – Consistency of Least Squares

Important

Theorem 5: Consistency of Least Squares

Under Assumption 1, $\hat{\mathbf{Q}}_{XX} \xrightarrow{p} \mathbf{Q}_{XX}$,

$$\hat{\mathbf{Q}}_{XY} \xrightarrow{p} \mathbf{Q}_{XY}, \quad \mathbf{Q}_{XX}^{-1} \xrightarrow{p} \mathbf{Q}_{XX}^{-1}, \quad \hat{\mathbf{Q}}_{Xe} \xrightarrow{p} 0,$$

and

$$\hat{\beta} \xrightarrow{p} \beta \quad \text{as } n \to \infty.$$

- **Theorem 5** states that $\hat{\beta}$ **converges in probability** to β as n increases – i.e., $\hat{\beta}$ is **consistent** for β.

Asymptotic Normality

- We have already shown that $\hat{\beta}$ **converges in probability** to β.
- **Consistency** is a good first step, but it does not describe the **distribution** of the estimator.
- Now, we derive the **asymptotic distribution** of $\hat{\beta}$.

Derivation of Asymptotic Distribution

- The derivation starts by writing the estimator as a function of sample moments.
- One moment must be normalized so that the **CLT** can be applied.
- Multiplying (17) by $\sqrt{n}$:

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i e_i \right) \tag{18}$$

- $\sqrt{n}(\hat{\beta} - \beta)$ is a function of:
  - The sample average $n^{-1} \sum_{i=1}^{n} X_i X_i'$.
  - The normalized sample average $n^{-1/2} \sum_{i=1}^{n} X_i e_i$.

Covariance Structure

- The random pairs (Y_i, X_i) are i.i.d., so $e_i = Y_i - X_i'\beta$ and $X_i e_i$ are also i.i.d.
- The product $X_i e_i$ is **mean-zero** $(E[Xe] = 0)$ with $k \times k$ **covariance matrix**:

$$\Omega = E\big[(Xe)(Xe)'\big] = E[XX'e^2].$$

Asymptotic Distribution

- We show below that Ω **has finite elements** under Assumption 2 (strengthened).
- Since $X_i e_i$ is i.i.d., mean zero, and has finite variance, the **CLT** implies:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i e_i \xrightarrow{d} N(0, \Omega) \tag{19}$$

Assumption 5 – Strengthened

Assumption 6 (Strengthened Assumption 5)

1. The variables (Y_i, X_i), $i = 1, \ldots, n$, are i.i.d.
2. $E[Y^4] < \infty$.
3. $E[\|X\|^4] < \infty$.
4. $\mathbf{Q}_{XX} = E[XX']$ is positive definite.

Note: For $X = (X_1, \ldots, X_k)'$, the fourth moment of the norm is $E[\|X\|^4] = E[(X'X)^2] = E\left[\left(\sum_{j=1}^k X_j^2\right)^2\right]$, which expands to $\sum_{j=1}^k E[X_j^4] + 2\sum_{j<l} E[X_j^2 X_l^2] < \infty$, requiring finite fourth moments for each regressor and their cross-products.

Theorem 6 – Asymptotic Normality

Important

Theorem 6: Asymptotic Normality
Assumption 2 (strengthened) implies that

$$\Omega < \infty \tag{20}$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i e_i \xrightarrow{d} N(0, \Omega) \tag{21}$$

as $n \to \infty$.

Asymptotic Distribution of $\hat{\beta}$

- Combining (15), (18), and (21):

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathbf{Q}_{XX}^{-1} N(0, \Omega) = N(0, \mathbf{Q}_{XX}^{-1} \Omega \mathbf{Q}_{XX}^{-1}).$$

- The final equality follows from the property that **linear combinations of normal vectors** are also **normal**.

Theorem 7 – Asymptotic Normality of Least Squares

Important

Theorem 7: Asymptotic Normality of Least Squares Estimator
Under Assumption 2 (strengthened), as $n \to \infty$:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \mathbf{V}_\beta) \quad (22)$$

where $\mathbf{Q}_{XX} = E[XX']$, $\Omega = E[XX'e^2]$, and

$$\mathbf{V}_\beta = \mathbf{Q}_{XX}^{-1} \Omega \mathbf{Q}_{XX}^{-1}.$$

Interpretation of Theorem 7

- The matrix $\mathbf{V}_\beta = \mathbf{Q}_{XX}^{-1}\Omega\mathbf{Q}_{XX}^{-1}$ is the **variance of the asymptotic distribution** of $\sqrt{n}(\hat{\beta} - \beta)$.
- This is often called the **asymptotic covariance matrix** of $\hat{\beta}$.
- The expression is known as the **sandwich form**, since Ω is "sandwiched" between two copies of $\mathbf{Q}_{XX}^{-1}$.

Comparing Asymptotic and Finite-Sample Variances

- It is useful to compare $\mathbf{V}_\beta$ from (22) with the finite-sample variance from (10):

$$\mathbf{V}_{\hat{\beta}} = \mathrm{var}[\hat{\beta} \mid \mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{D}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}.$$

- **Key observation**:
  - $\mathbf{V}_{\hat{\beta}}$ is the **exact conditional variance** of $\hat{\beta}$.
  - $\mathbf{V}_\beta$ is the **asymptotic variance** of $\sqrt{n}(\hat{\beta} - \beta)$.
  - Thus $\mathbf{V}_\beta$ should be roughly n times as large as $\mathbf{V}_{\hat{\beta}}$:

$$\mathbf{V}_\beta \approx n\mathbf{V}_{\hat{\beta}}.$$

Comparing Asymptotic and Finite-Sample Variances (Cont.)

- From (22): $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \mathbf{V}_\beta)$.
- Rearranging:

$$\hat{\beta} - \beta \approx \frac{1}{\sqrt{n}} N(0, \mathbf{V}_\beta).$$

- Taking variances on both sides:

$$\mathrm{Var}[\hat{\beta}] \approx \frac{1}{n} \mathbf{V}_\beta.$$

- Since $\mathbf{V}_{\hat{\beta}} = \mathrm{Var}[\hat{\beta} \mid \mathbf{X}]$ describes the same variance behavior in finite samples:

$$\mathbf{V}_{\hat{\beta}} \approx \frac{1}{n} \mathbf{V}_\beta, \quad \text{i.e.,} \quad \mathbf{V}_\beta \approx n \mathbf{V}_{\hat{\beta}}.$$

Estimation of the Asymptotic Covariance Matrix

- The asymptotic covariance matrix $\mathbf{V}_\beta = \mathbf{Q}_{XX}^{-1}\Omega\mathbf{Q}_{XX}^{-1}$ is unknown – it depends on population moments.
- A consistent estimator replaces each population moment with its sample analogue:

$$\hat{\mathbf{V}}_{\hat{\beta}} = \hat{\mathbf{Q}}_{XX}^{-1}\hat{\Omega}\hat{\mathbf{Q}}_{XX}^{-1}$$

where $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \hat{e}_i^2$ and $\hat{e}_i = Y_i - X_i'\hat{\beta}$.

Three Variance Concepts

- Three related but distinct objects appear in OLS inference:

Object	Definition	Role
$\mathbf{V}_\beta$	$\mathbf{Q}_{XX}^{-1}\Omega\mathbf{Q}_{XX}^{-1}$	Variance of limiting distribution of $\sqrt{n}(\hat{\beta}-\beta)$
$\mathbf{V}_{\hat{\beta}}$	$(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{D}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$	Exact conditional variance of $\hat{\beta}$ for given n
$\hat{\mathbf{V}}_{\hat{\beta}}$	$\hat{\mathbf{Q}}_{XX}^{-1}\hat{\Omega}\hat{\mathbf{Q}}_{XX}^{-1}$	Feasible estimator computed from data

Three Variance Concepts (Cont.)

- $\mathbf{V}_\beta$ and $\mathbf{V}_{\hat{\beta}}$ are both **population quantities** – never observed, depend on unknown moments.
- They are related by $\mathbf{V}_\beta \approx n\mathbf{V}_{\hat{\beta}}$ in large samples.
- $\hat{\mathbf{V}}_{\hat{\beta}}$ is what you **actually compute**:

$$\hat{\mathbf{V}}_{\hat{\beta}} = \hat{\mathbf{Q}}_{XX}^{-1}\hat{\Omega}\hat{\mathbf{Q}}_{XX}^{-1}, \qquad \hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \hat{e}_i^2$$

- It is **consistent** for both $\frac{1}{n}\mathbf{V}_\beta$ and $\mathbf{V}_{\hat{\beta}}$ as $n \to \infty$.
- The square roots of its diagonal entries are the **standard errors** used in t-tests and the Wald test.

Hypothesis Testing under the Large Sample Properties

Large Sample t-Test

Important

Large Sample t-Statistic

To test $H_0 : \beta_j = \beta_{j,0}$, use

$$t_j = \frac{\hat{\beta}_j - \beta_{j,0}}{\widehat{\mathrm{se}}(\hat{\beta}_j)}$$

where $\widehat{\mathrm{se}}(\hat{\beta}_j)$ is a consistent estimator of the standard error.

- Under H_0 and regularity conditions:

$$t_j \xrightarrow{d} \mathcal{N}(0, 1)$$

Decision Rule – Large Sample t-Test

- For a two-sided test at level α, reject H_0 if:

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

- For a one-sided test:
  - Right tail: reject if $t_j > z_{1-\alpha}$
  - Left tail: reject if $t_j < -z_{1-\alpha}$
- z_q is the q-quantile of the standard normal distribution.

Joint Tests on β – The Linear Case

- **Setup:** test q linear restrictions jointly

$$H_0: \; R\beta = r$$

where R is $q \times k$ and r is $q \times 1$, both constants

- **Example** in a regression with $\beta_1, \beta_2, \beta_3, \beta_4$:

$$H_0: \; \beta_1 = \beta_2 \quad \text{and} \quad \beta_3 + \beta_4 = 1$$

- Stack into $R\beta = r$ form:

$$R = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \qquad r = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

- **Wald statistic:**

$$W = (R\hat{\beta} - r)^\top \left[R \,\widehat{\text{Var}}(\hat{\beta})\, R^\top\right]^{-1} (R\hat{\beta} - r) \xrightarrow{d} \chi^2_q$$

- Under homoskedasticity, equivalently $F = W/q \sim F_{q,\, n-k}$

Required Reading

- **Stock & Watson**, *Introduction to Econometrics*, 4th Edition
  - Chapter 6, Sections 6.5–6.6 (pp. 225–229) – least squares assumptions and distribution of OLS in multiple regression.
  - Chapter 7, Sections 7.1–7.2 (pp. 247–256) – hypothesis tests and confidence intervals, joint F-test.
  - Appendix 6.2 – distribution of OLS estimators under homoskedasticity.
  - Appendix 7.1 – the Bonferroni test of a joint hypothesis.

- **Wooldridge**, *Introductory Econometrics*, 7th Edition
  - Advanced Treatment E, Section E.3 (pp. 767–769) – statistical inference under normality; proof that the t-statistic follows t_{n-k-1} (Theorems E.5–E.6).
  - Advanced Treatment E, Section E.4 (pp. 769–771) – asymptotic analysis and Wald statistics for joint hypotheses.