Introduction to Econometrics & OLS

ECON0004: Applied Econometrics

Ambrose Wang

University College London

March 2026

Based on: Wooldridge (2019) *Introductory Econometrics*, 7e

Outline

What Is Econometrics?

Definition: Use of statistical and mathematical methods to give empirical content to economic relationships.

Four core uses:

1. Estimate structural parameters of economic models
   e.g. price elasticity, returns to education
2. Test economic theories
3. Forecast economic variables
4. Evaluate and implement policies

> **Key challenge:** economists rarely have experimental data.
>
> We rely on **observational** (non-experimental) data and must account for confounding.

Our motivating question:

Does more schooling raise wages?

We will answer this with real data by the end of the lecture.

Types of Economic Data

Cross-sectional data

Observations on *different* units at a *single* point in time.

E.g. wages of 1,822 workers surveyed in 2020.

- Order of observations does not matter
- Variables can be binary (married/not)

Time series data

Observations on a *single* unit over *many* periods.

E.g. quarterly UK GDP, 1980–2023.

$\Rightarrow$ Observations rarely independent (serial dependence).

Pooled cross-sections

Multiple cross-sections from *different* periods combined.

Different individuals surveyed each year.

Panel (longitudinal) data

The *same* units observed over *multiple* periods.

Track the same firms annually.

> In this course we focus on **cross-sectional data** with **random sampling** (SLR.2).

Causality vs. Correlation

Correlation $\neq$ Causation $\quad r = \dfrac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(n-1)\, s_x s_y}$

Spurious correlation arises from:

1. **Confounding / omitted variables** — a third variable affects both x and y (e.g. ability drives both wages and schooling choice)
2. **Reverse causality** — we cannot determine which variable causes which
3. **Indirect causation** — $A \to B \to C$ but we observe only $A \sim C$
4. Purely coincidental association

> Observing a correlation between schooling and wages does **not** imply that education *causes* higher wages. Ability, family background, etc. may drive both.

Ceteris Paribus & Counterfactual Reasoning

Ceteris paribus ("other things equal"):

- Economists want to isolate the effect of *one* variable, holding everything else constant.
- In practice, it is **impossible** to literally hold everything fixed.

> The key question is:
>
> Have *enough* other factors been held fixed to make a **causal claim**?

Counterfactual reasoning:

Imagine two otherwise identical individuals who differ *only* in education level.

The wage gap between them measures the **causal effect** of schooling.

$\Rightarrow$ Regression attempts to approximate this comparison using observational data.

Steps in Empirical Economic Analysis

1. **Economic model**
   Mathematical equations describing relationships.
   "How, and to what extent, does one thing affect another?"
2. **Econometric model**
   Statistical model specifying the relationship between variables.
   Requires: (i) functional form, (ii) **error term** (unobserved factors, also called the disturbance), (iii) parameters to be estimated.
3. **Collect data**
   Cross-sectional sample, survey, administrative records, ...
4. **Estimate parameters & test hypotheses**
   Apply OLS (or other estimators); draw statistical inference.

> The error term **is not a modelling failure** — it captures genuine randomness and unobserved heterogeneity inherent in human behaviour.

Running Example: Does Schooling Raise Wages?

Variables:

- $\ln w_i$ = log hourly wage of individual i
- edu_i = years of schooling of individual i

Why log wage?

For small changes in edu:

$$\Delta \ln w_i \approx \frac{\Delta w_i}{w_i} = \%\Delta w_i$$

So β_1 is a **semi-elasticity**: one extra year of schooling $\approx 100\beta_1\%$ wage increase.

Population model:

$$\ln w_i = \beta_0 + \beta_1 edu_i + u_i$$

- $\beta_0 > 0$: log wage at zero education
- $\beta_1 > 0$: return to schooling
- u_i: ability, family background, luck, ...

Preview: using $n = 1{,}822$ workers:
$\hat{\beta}_1 = 0.0658$
each extra year $\approx +6.58\%$ wages

The Population Model

Simple Linear Regression (SLR) Model

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \qquad i = 1, \ldots, n$$

- β_0 — **intercept**: expected value of Y when $X = 0$
- β_1 — **slope**: change in Y per unit increase in X, *ceteris paribus*
- u_i — **error term**: all factors affecting Y other than X

β_0 and β_1 are **unknown population parameters**.

They describe the true relationship in the population — our goal is to estimate them from data.

u_i captures **everything not in the model**:

ability, family background, luck, measurement error, ...

The error makes exact prediction impossible, but OLS estimates the *average* relationship.

Deterministic vs. Stochastic Relationship

Deterministic (unrealistic):

$$\ln w_i = \alpha + \beta \cdot edu_i$$

- Perfect linear relationship
- Every observation lies *exactly* on the line
- Impossible in practice — human behaviour is noisy

Stochastic (realistic):

$$\ln w_i = \beta_0 + \beta_1 edu_i + u_i$$

- Observations *scattered* around the line
- u_i captures unobserved factors (ability, etc.)
- Goal: estimate β_0, β_1 from data

The error term u_i is **not** a modelling failure. It reflects the genuine randomness and unobserved heterogeneity that make exact deterministic relationships impossible in economics.

Population Regression Function (PRF)

Take the conditional expectation of Y_i given X_i:

$$\begin{aligned}\mathbb{E}[Y_i \mid X_i] &= \mathbb{E}[\beta_0 + \beta_1 X_i + u_i \mid X_i] \\ &= \beta_0 + \beta_1 X_i + \underbrace{\mathbb{E}[u_i \mid X_i]}_{=0 \text{ (by SLR.4)}}\end{aligned}$$

$$\boxed{\mathbb{E}[Y_i \mid X_i] = \beta_0 + \beta_1 X_i}$$

Population Regression Function

$\mathbb{E}[Y \mid X] = \beta_0 + \beta_1 X$ is the **conditional mean** of Y given X.

It traces the expected value of Y for each value of X in the population.

Key insight:

β_1 measures how the *conditional mean* of Y changes with X.

This is exactly what OLS estimates from the sample.

The zero conditional mean assumption (SLR.4) is what makes this interpretation

Fitted Values, Residuals, and Errors

Fitted (predicted) value:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

OLS residual:

$$\hat{u}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

Error u_i (population):
unobservable; true deviation from the PRF

Residual $\hat{u}_i$ (sample):
computable; deviation from the *estimated* line

Example

Worker i has $edu_i = 12$.

Fitted: $\hat{y}_i = 0.894 + 0.0658 \times 12 = 1.683$

Actual: $\ln w_i = 1.90$

Residual: $\hat{u}_i = 1.90 - 1.683 = 0.217$

This worker earns more than education alone predicts — perhaps high unobserved ability.

Why Minimise *Squared* Residuals?

Goal: choose $\hat{\beta}_0$, $\hat{\beta}_1$ so the fitted line "best fits the data."

Why not minimise $\sum |\hat{u}_i|$ (absolute residuals)?
$\Rightarrow$ Not differentiable at zero; no unique closed-form solution.

OLS Criterion: Minimise SSR

$$\text{SSR}(\beta_0, \beta_1) = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

Advantages of squaring:

- Penalises large errors more heavily than small ones
- Differentiable everywhere $\Rightarrow$ calculus applies
- Unique closed-form solution exists (under SLR.3)
- Directly related to variance: minimising SSR $\approx$ minimising unexplained noise

Two Key Properties of the OLS Fitted Line

These follow directly from the first-order conditions (derived in the next section):

Property 1:
The line passes through the sample means $(\bar{X}, \bar{Y})$:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Regardless of the data, the OLS line always passes through $(\bar{X}, \bar{Y})$.

Property 2:
OLS residuals are orthogonal to the data:

$$\sum_{i=1}^{n} \hat{u}_i = 0, \qquad \sum_{i=1}^{n} \hat{u}_i X_i = 0$$

$\Rightarrow$ The average residual is zero; residuals and X are uncorrelated in the sample.

> These are **algebraic facts**, not statistical assumptions. They hold in every sample, regardless of whether the model is correctly specified.

OLS: The Minimisation Problem

We want to find $\hat{\beta}_0$, $\hat{\beta}_1$ that solve:

$$\min_{\beta_0, \beta_1} \text{SSR}(\beta_0, \beta_1) = \min_{\beta_0, \beta_1} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

Approach: take partial derivatives (first-order conditions) and set to zero.

Two unknowns:

- β_0 (intercept)
- β_1 (slope)

Two first-order conditions (FOCs):

$$\frac{\partial \text{SSR}}{\partial \beta_0} = 0$$

$$\frac{\partial \text{SSR}}{\partial \beta_1} = 0$$

Solving both FOCs simultaneously gives the **OLS estimators** $\hat{\beta}_0$ and $\hat{\beta}_1$.

FOC 1: Differentiating w.r.t. β_0

Step 1. Differentiate $\mathrm{SSR} = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$ w.r.t. β_0 and set to zero:

$$\frac{\partial \,\mathrm{SSR}}{\partial \beta_0} = -2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i) = 0$$

Step 2. Divide by -2 and distribute the summation:

$$\sum_{i=1}^{n} Y_i \;=\; n\beta_0 + \beta_1 \sum_{i=1}^{n} X_i \quad \textbf{(Normal Equation 1)}$$

Note: $\sum_{i=1}^{n} \beta_0 = n\beta_0$ since β_0 is a constant.

Solving for $\hat{\beta}_0$

Starting from Normal Equation 1: $\sum_{i=1}^{n} Y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} X_i$

Step 1. Divide both sides by n:

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} Y_i}_{\bar{Y}} = \beta_0 + \beta_1 \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\bar{X}} \quad \Longrightarrow \quad \bar{Y} = \beta_0 + \beta_1 \bar{X}$$

Step 2. Solve for $\hat{\beta}_0$:

$$\boxed{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}}$$

The OLS fitted line **always passes through** $(\bar{X}, \bar{Y})$.

FOC 2: Differentiating w.r.t. β_1

Step 1. Differentiate SSR w.r.t. β_1 and set to zero:

$$\frac{\partial \,\mathrm{SSR}}{\partial \beta_1} = -2\sum_{i=1}^{n} X_i(Y_i - \beta_0 - \beta_1 X_i) = 0$$

Rearranging — **Normal Equation 2:**

$$\sum_{i=1}^{n} X_i Y_i = \beta_0 \sum_{i=1}^{n} X_i + \beta_1 \sum_{i=1}^{n} X_i^2$$

We now have two normal equations (Slides 16 & 17).
Next: substitute $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ into Normal Equation 2 to isolate $\hat{\beta}_1$.

FOC 2 Continued: Isolating $\hat{\beta}_1$

Substitute $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ and use $\sum X_i = n\bar{X}$ into Normal Equation 2:

Step 1.

$$\begin{aligned}\sum X_i Y_i &= (\bar{Y} - \hat{\beta}_1 \bar{X})\, n\bar{X} + \hat{\beta}_1 \sum X_i^2 \\ &= n\bar{X}\bar{Y} + \hat{\beta}_1 \left(\sum X_i^2 - n\bar{X}^2 \right)\end{aligned}$$

Rearranging:

$$\sum X_i Y_i - n\bar{X}\bar{Y} = \hat{\beta}_1 \left(\sum X_i^2 - n\bar{X}^2 \right)$$

Let $A = \sum X_i Y_i - n\bar{X}\bar{Y}$ and $B = \sum X_i^2 - n\bar{X}^2$.
Then $\hat{\beta}_1 = A/B$. The next two slides show that $A = \sum (X_i - \bar{X})(Y_i - \bar{Y})$ and $B = \sum (X_i - \bar{X})^2$.

Derivation: Expanding the Numerator

Need to simplify: $A = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}$

Step 1. Write $X_i = (X_i - \bar{X}) + \bar{X}$ and $Y_i = (Y_i - \bar{Y}) + \bar{Y}$, then expand:

$$\sum X_i Y_i = \sum \big[(X_i - \bar{X})(Y_i - \bar{Y}) + \bar{Y}(X_i - \bar{X}) + \bar{X}(Y_i - \bar{Y}) + \bar{X}\bar{Y}\big]$$

Step 2. Use the fact that $\sum (X_i - \bar{X}) = 0$ and $\sum (Y_i - \bar{Y}) = 0$:

$$\sum X_i Y_i = \sum (X_i - \bar{X})(Y_i - \bar{Y}) + \bar{Y} \cdot 0 + \bar{X} \cdot 0 + n\bar{X}\bar{Y}$$

$$\therefore \quad A = \sum X_i Y_i - n\bar{X}\bar{Y} = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

Derivation: Expanding the Denominator

Need to simplify: $B = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$

Step 1. Write $X_i = (X_i - \bar{X}) + \bar{X}$ and expand X_i^2:

$$\begin{aligned}\sum X_i^2 &= \sum\left[(X_i - \bar{X})^2 + 2\bar{X}(X_i - \bar{X}) + \bar{X}^2\right] \\ &= \sum (X_i - \bar{X})^2 + 2\bar{X}\underbrace{\sum (X_i - \bar{X})}_{=0} + n\bar{X}^2\end{aligned}$$

Step 2. Use $\sum (X_i - \bar{X}) = 0$:

$$\sum X_i^2 = \sum (X_i - \bar{X})^2 + n\bar{X}^2$$

$$\therefore \quad B = \sum X_i^2 - n\bar{X}^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

OLS Estimators: The Closed-Form Solution

Substituting A and B into $A = \hat{\beta}_1 B$ from Slide 17:

OLS Estimators

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\widehat{\text{Cov}}(X, Y)}{\widehat{\text{Var}}(X)} \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$\hat{\beta}_1$ = **sample covariance** of X and Y, divided by **sample variance** of X.

The sign of $\hat{\beta}_1$ equals the sign of $\widehat{\text{Cov}}(X, Y)$.

Wage example:

- $\hat{\beta}_1 = 0.0658$
- $\hat{\beta}_0 = 0.894$
- $\hat{Y}_i = 0.894 + 0.0658\, edu_i$

Desirable Properties of Estimators

OLS gives us $\hat{\beta}_1$ — but how good is it as an estimator of the true β_1?

Unbiasedness

$\mathbb{E}[\hat{\beta}] = \beta$

On average across repeated samples, the estimator recovers the true value.

$$\text{Bias}(\hat{\beta}) = \mathbb{E}[\hat{\beta}] - \beta = 0$$

Efficiency

Smallest $\text{Var}(\hat{\beta})$ among all unbiased estimators.

Multiple unbiased estimators may exist — prefer the one with lowest variance.

$$\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$$

Consistency

$\hat{\beta} \to_p \beta$
as $n \to \infty$

As sample size grows, the estimator converges to the true value in probability.

Unbiased $\neq \hat{\beta} = \beta$ in any given sample. It is a statement about the *distribution* of $\hat{\beta}$ across repeated samples — not about a single estimate.

SLR Assumptions: Overview

Under six assumptions, OLS has strong guarantees.

	Assumption	Content	What it delivers
SLR.1	Linearity	Model linear in β	Model is identified
SLR.2	Random sampling	i.i.d. observations	Inference is valid
SLR.3	Variation in X	$\sum(X_i - \bar{X})^2 > 0$	$\hat{\beta}_1$ exists
SLR.4	Zero conditional mean	$\mathbb{E}[u\|X] = 0$	**Unbiasedness**
SLR.5	Homoskedasticity	$\text{Var}(u\|X) = \sigma^2$	Gauss–Markov (BLUE)
SLR.6	Normality	$u \sim \mathcal{N}(0, \sigma^2)$	Exact t-tests

SLR.4 is the most critical assumption. If it fails (e.g. omitted variable), OLS is **biased and inconsistent**.

SLR.1: Linearity in Parameters

SLR.1 — Linearity in Parameters

The population model is linear in its parameters β_0 and β_1:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

What "linear" means:

Linearity refers to the **parameters** β_0, β_1, **not** necessarily to the variables.

Non-linear transformations of X are perfectly allowed:

- $\ln X_i$, X_i^2, $\sqrt{X_i}$ are all valid regressors

Example

Allowed: $Y_i = \beta_0 + \beta_1 \ln X_i + u_i$
($\ln X_i$ enters linearly in β_1)

NOT allowed: $Y_i = \beta_0 + \beta_1^2 X_i + u_i$
(parameter β_1 enters non-linearly)

SLR.2: Random Sampling

SLR.2 — Random Sampling

The data $\{(X_i, Y_i)\}_{i=1}^n$ are an **i.i.d.** (independent and identically distributed) random sample from the population.

What this means:

(X_i, u_i) is independent of (X_j, u_j) for $i \neq j$. Each observation is an independent draw from the same population distribution.

When it fails:

- Time series data (serial dependence)
- Cluster effects (e.g. students in same class)

Example

Our data: 1,822 workers drawn randomly from the NLSY survey.

Each worker's wage and education is independent of every other worker's — SLR.2 is plausible here.

SLR.3: Sample Variation in X

SLR.3 — Sample Variation in X

The sample values of X_i are not all the same:

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 > 0$$

Why it is needed:

The OLS estimator is:

$$\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

If all X_i are identical, the denominator equals zero — $\hat{\beta}_1$ is **undefined**.

Example

Example of failure:
Suppose all 1,822 workers have exactly 12 years of education.

We observe many wage values at a single X point — the slope is unidentifiable.

SLR.4: Zero Conditional Mean — The Critical Assumption

SLR.4 — Zero Conditional Mean (Exogeneity)

$$\mathbb{E}[u_i \mid X_i] = 0$$

The error term has **zero expected value** at every value of X. X carries *no information* about the mean of u.

When does SLR.4 fail?

- **Omitted variable** correlated with X
  ability influences both wages and schooling decisions $\Rightarrow \text{Cov}(u, edu) \neq 0$
- **Reverse causality** (Y affects X in addition to X affecting Y)
- Measurement error in X (recorded schooling differs from true schooling)

SLR.4 is the **key identification assumption**. Without it, OLS is biased. Randomised experiments guarantee $\mathbb{E}[u|X] = 0$ *by design* — this is why they are the gold standard for causal inference.

SLR.5: Homoskedasticity

SLR.5 — Homoskedasticity

The variance of the error is **constant** for all values of X:

$$\text{Var}(u_i \mid X_i) = \sigma^2 \quad \text{(a constant)}$$

Homoskedastic (SLR.5 holds):

Scatter around the regression line is equally spread at all values of X.

Heteroskedastic (SLR.5 violated):

Spread of errors varies with X.

E.g. variance of wages may increase with education — higher earners face more idiosyncratic variation.

Why it matters:

- Needed for OLS to be **BLUE** (minimum variance among linear unbiased estimators)
- Without it, OLS standard errors may be misestimated
- In practice: use *heteroskedasticity-robust* standard errors (Lecture 4 onwards)

SLR.6: Normality of Errors

SLR.6 — Normality

The population error is independent of X and normally distributed:

$$u \sim \mathcal{N}(0, \sigma^2)$$

Why it is needed:

Under SLR.6, $\hat{\beta}_1$ is *exactly* normally distributed in **finite samples**, enabling exact t-tests.

When can we relax it?

For large n, the Central Limit Theorem ensures $\hat{\beta}_1$ is *approximately* normal without SLR.6.

t-statistics remain valid asymptotically.

SLR.6 is the **least critical** of the six assumptions for applied inference with large samples.

In our wage example:
$n = 1{,}822$ — CLT applies, normality is not required.

Unbiasedness of OLS: Proof

Claim: Under SLR.1–SLR.4, $\mathbb{E}[\hat{\beta}_1] = \beta_1$.

Step 1. Substitute $Y_i = \beta_0 + \beta_1 X_i + u_i$ into $\hat{\beta}_1 = \sum(X_i - \bar{X})Y_i / \sum(X_i - \bar{X})^2$. Using $\sum(X_i - \bar{X}) = 0$, the β_0 term vanishes:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(X_i - \bar{X})\, u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Step 2. Take conditional expectation given $X_1, \ldots, X_n$. By **SLR.4**: $\mathbb{E}[u_i \mid X_i] = 0$, so each term $\mathbb{E}[(X_i - \bar{X})u_i \mid X_1, \ldots, X_n] = 0$:

$$\mathbb{E}\left[\hat{\beta}_1 \,\middle|\, X_1, \ldots, X_n\right] = \beta_1 + 0 = \beta_1$$

Taking unconditional expectation: $\mathbb{E}[\hat{\beta}_1] = \beta_1$. $\square$

Gauss–Markov Theorem: OLS Is BLUE

Gauss–Markov Theorem

Under **SLR.1–SLR.5**, the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are the **Best Linear Unbiased Estimators (BLUE)**.

Unpacking "BLUE":

- **B**est: minimum variance
- **L**inear: estimator is a linear function of $\{Y_i\}$
- **U**nbiased: $\mathbb{E}[\hat{\beta}_j] = \beta_j$
- **E**stimator: constructed from the observed data

Consequence:

Among all estimators that are linear and unbiased, OLS has the *smallest sampling variance*. No other linear unbiased estimator does better.

Gauss–Markov requires **SLR.5** (homoskedasticity).

With heteroskedastic errors, *Weighted OLS* (WLS) can be more efficient.

Omitted Variable Bias (OVB): Preview

Suppose the **true** model includes x_2 which is omitted:

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

We estimate $Y_i = \beta_0 + \beta_1 X_{1i} + v_i$, where $v_i = u_i + \beta_2 X_{2i}$. If $\text{Cov}(X_2, X_1) \neq 0$, then $\mathbb{E}[v_i | X_{1i}] \neq 0$ — **SLR.4 fails**.

OVB Formula:

$$\text{Bias}[\hat{\beta}_1] = \frac{\beta_2 \ \text{Cov}(X_2,\ X_1)}{\text{Var}(X_1)}$$

Direction of bias:

	$\text{Cov} > 0$	$\text{Cov} < 0$
$\beta_2 > 0$	+ **bias**	− **bias**
$\beta_2 < 0$	− **bias**	+ **bias**

Example

Wage regression: ability x_2 drives wages ($\beta_2 > 0$) and is correlated with education ($\text{Cov} > 0$) $\Rightarrow$ **upward bias** in $\hat{\beta}_1$.

Full treatment: Lecture 5.

What the Assumptions Buy Us: A Hierarchy

Assumptions	Result for OLS	Key requirement
SLR.1–SLR.2	Model identifiable; inference valid	Correct spec. + i.i.d. data
SLR.1–SLR.3	$\hat{\beta}_1$ *exists* (denom. $\neq 0$)	Variation in X
SLR.1–SLR.4	OLS is **unbiased**: $\mathbb{E}[\hat{\beta}_1] = \beta_1$	$\mathbb{E}[u \mid X] = 0$
SLR.1–SLR.5	OLS is **BLUE** (minimum variance)	Homoskedasticity
SLR.1–SLR.6	Exact t-tests in **finite samples**	Normality of errors

The hierarchy: each assumption adds a new guarantee. SLR.4 (unbiasedness) is the most critical for applied work. SLR.6 (normality) is the least critical with large n (CLT takes over).

Decomposing Total Variation: $\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR}$

Start from $Y_i = \hat{Y}_i + \hat{u}_i$. Subtract $\bar{Y}$: $Y_i - \bar{Y} = (\hat{Y}_i - \bar{Y}) + \hat{u}_i$. Square and sum:

$$\underbrace{\sum (Y_i - \bar{Y})^2}_{\mathrm{SST}} = \underbrace{\sum (\hat{Y}_i - \bar{Y})^2}_{\mathrm{SSE}} + \underbrace{\sum \hat{u}_i^2}_{\mathrm{SSR}} + 2\underbrace{\sum (\hat{Y}_i - \bar{Y})\hat{u}_i}_{=0}$$

The cross-product is zero because $\sum \hat{u}_i \hat{Y}_i = 0$ (from the normal equations).

$$\boxed{\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR}}$$

Term	Formula	Meaning
SST	$\sum (Y_i - \bar{Y})^2$	Total variation in Y
SSE	$\sum (\hat{Y}_i - \bar{Y})^2$	Variation *explained* by the model
SSR	$\sum \hat{u}_i^2$	Variation *unexplained* (residuals)

R^2: Coefficient of Determination

R^2 (R-squared)

$$R^2 = 1 - \frac{\text{SSR}}{\text{SST}} = \frac{\text{SSE}}{\text{SST}}$$

The proportion of total variation in Y **explained by the model**.

Properties:

- $0 \leq R^2 \leq 1$
- $R^2 = 1$: all data on the regression line (SSR $= 0$)
- $R^2 = 0$: model explains nothing beyond $\bar{Y}$
- In SLR: $R^2 = r_{XY}^2$ (squared sample correlation)

Caveats:

- **No universal benchmark** for "high" or "low"
- Sensitive to functional form (log vs. level)
- Adding regressors always raises R^2 — use adjusted $\bar{R}^2$ instead

Wage example: $R^2 = 0.1231$
Education explains 12.3% of wage variation.

Standard Error of the Equation (SEE)

A low SSR alone does not tell us whether the fit is "good" — it depends on the scale of Y.

Standard Error of the Equation (SEE)

$$\text{SEE} = \hat{\sigma} = \sqrt{\frac{\text{SSR}}{n-K}}$$

where n = sample size, K = number of parameters estimated ($\hat{\beta}_0$ + slopes).

In SLR: $K = 2$, so:

$$\hat{\sigma} = \sqrt{\frac{\text{SSR}}{n-2}}$$

The degrees-of-freedom correction ensures $\hat{\sigma}^2$ is an **unbiased** estimator of $\sigma^2 = \text{Var}(u)$.

Wage example:

Stata reports: Root MSE = 0.4436

Predicted log wages deviate from actual by ± 0.44 on average.

> SEE is **scale-dependent** — unlike R^2, cannot be compared across models with different Y units.

Variance of the OLS Slope Estimator

Under SLR.1–SLR.5:

Variance and Standard Error of $\hat{\beta}_1$

$$\mathrm{Var}\left(\hat{\beta}_1 \,\middle|\, X_1, \ldots, X_n\right) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad \mathrm{se}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

What increases $\mathrm{se}(\hat{\beta}_1)$?

- $\uparrow \sigma^2$: noisier data $\Rightarrow$ less precise estimates
- $\downarrow \sum(X_i - \bar{X})^2$: less variation in $X \Rightarrow$ harder to pin down the slope
- $\downarrow n$: smaller sample $\Rightarrow$ less information

Wage example: $\mathrm{se}(\hat{\beta}_1) = 0.00412$ — very precise due to large $n = 1{,}822$ and substantial variation in education.

Hypothesis Testing: The t-Statistic

Standard null: $H_0 : \beta_1 = 0$ (education has no effect on wages).

t-Statistic

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - \beta_1^0}{\text{se}(\hat{\beta}_1)} \sim t_{n-K} \quad (= t_{n-2} \text{ in SLR, since } K = 2)$$

Under $H_0 : \beta_1 = 0$: $t_{\hat{\beta}_1} = \dfrac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)}$

Two-sided test ($H_1 : \beta_1 \neq 0$):

- Reject H_0 if $|t| > c_{\alpha/2}$
- At 5% level, large n: $c_{0.025} \approx 1.96$

One-sided test ($H_1 : \beta_1 > 0$):

- Reject H_0 if $t > c_\alpha$

Example

Wage example:

$$t = \frac{0.0658}{0.00412} = 15.99$$

$|15.99| \gg 1.96 \Rightarrow$ **Reject H_0**

Education is **highly statistically significant**.

Confidence Intervals

$(1-\alpha)\times 100\%$ Confidence Interval for β_1

$$\hat{\beta}_1 \pm t_{n-K,\,\alpha/2} \times \text{se}(\hat{\beta}_1)$$

The set of all values b for which $H_0 : \beta_1 = b$ is **not** rejected at significance level α.

Correct interpretation:

If we repeated sampling many times, 95% of constructed intervals would contain the true β_1.

Note: the CI is *random* (depends on the sample); the true β_1 is fixed — it is either in the interval or not.

Example

Wage example (95% CI):

$$0.0658 \pm 1.96 \times 0.00412$$

$$= [0.0577,\ \ 0.0739]$$

One extra year of schooling raises wages by between **5.8% and 7.4%**.

Reading Stata Output

Command: `reg lnw edu`

Source	SS	df	MS	
Model	50.30	1	50.30	$n = 1{,}822$
Residual	358.17	1820	0.197	$F(1, 1820) = 255.6$
Total	408.47	1821	0.224	Prob $> F = 0.000$
				$R^2 = 0.1231$
				Root MSE $= 0.4436$

lnw	Coef.	SE	t	$P>\|t\|$	[95% CI]
edu	**0.0658**	0.0041	**15.99**	**0.000**	[0.058, 0.074]
_cons	0.894	0.0769	11.63	0.000	[0.743, 1.045]

Key readings:

- $n = 1{,}822$ observations
- $\hat{\beta}_1 = 0.0658$:
  $+1$ year edu $\Rightarrow +6.58\%$ wages
- $\text{se}(\hat{\beta}_1) = 0.00412$
- $t = 15.99 \gg 1.96$, $p = 0.000$:
  reject $H_0 : \beta_1 = 0$ — highly significant
- $R^2 = 0.123$:
  edu explains 12.3% of wage variation

Each additional year of schooling is *associated with* a **6.58%** increase in hourly wages (correlation, not causation — SLR does not control for ability).

Lecture Summary

Pt. I–II: What Is Econometrics? & The SLR Model

- Econometrics: empirical analysis with observational data
- SLR model: $Y_i = \beta_0 + \beta_1 X_i + u_i$
- PRF: $\mathbb{E}[Y|X] = \beta_0 + \beta_1 X$ (under SLR.4)
- Error u_i vs. residual $\hat{u}_i$

Pt. III: OLS Derivation (SSR Minimisation)

- Minimise SSR via two FOCs
- $\hat{\beta}_1 = \mathrm{Cov}(X, Y)/\mathrm{Var}(X)$
- $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

Pt. IV: Assumptions & Properties (SLR.1–6)

- **SLR.4 most critical** — $\mathbb{E}[u|X] = 0$
- OLS unbiased under SLR.1–4
- OLS is BLUE under SLR.1–5
- OVB: $\mathrm{Bias}[\hat{\beta}_1] = \beta_2 \,\mathrm{Cov}(X_2, X_1)/\mathrm{Var}(X_1)$

Pt. V: Goodness of Fit & Inference

- $\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR}$; $R^2 = \mathrm{SSE}/\mathrm{SST} = 0.123$
- $\mathrm{Var}(\hat{\beta}_1) = \sigma^2 / \sum (X_i - \bar{X})^2$
- $t = 15.99$; 95% CI $= [0.058,\ 0.074]$

Next: Lecture 2 *— OLS assumptions in depth, unbiasedness proof, and the Gauss–Markov theorem.*