Econometrics

Lecture 2: Simple Regression

Recap: Foundations from Lecture 1

Key results we will build on:

- The **conditional expectation function** (CEF) $\mathbb{E}[Y \mid X]$ is the best predictor of Y given X
- The **law of iterated expectations** (LIE): $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$
- **Estimators are random variables** — properties depend on their sampling distributions
- The **central limit theorem** (CLT) justifies Normal-based inference for large samples

> Today's question: how do we *estimate* $\mathbb{E}[Y \mid X]$ from data, using a linear model?

How Does Y Relate to X?

Motivating questions:

- How does an additional year of **education** affect **wages**?
- How does **class size** affect student **test scores**?
- How does **fertilizer use** affect crop **yield**?
- How does **job training** affect worker **productivity**?

In each case, we want to:

1. **Quantify** the relationship between Y and X
2. **Test** whether the relationship is statistically significant
3. **Interpret** the result — ideally as a causal effect

The Simple Linear Regression Model

Simple Linear Regression

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

Components:

- Y_i: **dependent variable** (outcome, response, regressand)
- X_i: **independent variable** (regressor, covariate, explanatory variable)
- β_0: **intercept** parameter
- β_1: **slope** parameter
- u_i: **error term** (disturbance) — everything affecting Y besides X

“Simple” = one explanatory variable. Multiple regression (Lecture 3) allows many.

Interpreting β_0 and β_1

In the model $Y_i = \beta_0 + \beta_1 X_i + u_i$:

> A one-unit increase in X changes Y by β_1 units, holding u fixed. Under $\mathbb{E}[u \mid X] = 0$ (SLR.4, formalized shortly): $\partial\, \mathbb{E}[Y \mid X]/\partial X = \beta_1$.

Intercept β_0: predicted value of Y when $X = 0$

- Often not economically meaningful (is $X = 0$ realistic?)

Slope β_1: the marginal effect of X on Y

- ***Ceteris paribus*** interpretation requires assumptions on u
- "Holding other factors constant" — but which factors?

The Error Term u_i

The error term u_i represents **all factors** affecting Y_i besides X_i:

In the wage equation $\ln(\text{wage}_i) = \beta_0 + \beta_1\,\text{educ}_i + u_i$, the error includes:

- Innate ability and intelligence
- Family background and connections
- Work experience and tenure
- Motivation and work ethic
- Measurement error in wages or education

> The error term is **not** just “randomness.” It contains real, potentially important factors. Whether ordinary least squares (OLS) gives us a causal estimate depends on how u_i relates to X_i.

Population vs. Sample Regression

Population Regression Function (PRF)

$$\mathbb{E}[Y \mid X] = \beta_0 + \beta_1 X$$

- Describes the *true* relationship
- β_0, β_1 are unknown parameters
- We never observe this directly

Sample Regression Function (SRF)

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

- Our *estimate* from data
- $\hat{\beta}_0, \hat{\beta}_1$ computed from sample
- Changes with every new sample

The goal: use the SRF to learn about the PRF. How close is $\hat{\beta}_1$ to β_1?

Returns to Education Revisited

From Lecture 1, the **Mincer wage equation** (Mincer, 1974):

$$\ln(\text{wage}_i) = \beta_0 + \beta_1 \, \text{educ}_i + u_i$$

Interpretation:

- Because the dependent variable is $\ln(\text{wage})$, β_1 measures the **approximate percentage change** in wages per additional year of education
- If $\hat{\beta}_1 = 0.08$: one more year of education is associated with $\approx 8\%$ higher wages

The challenge:

- People with more education may also have higher ability
- If ability is in u_i and correlated with educ_i, then $\hat{\beta}_1$ captures *both* the effect of education *and* ability
- This is the **omitted variable bias** problem (Lecture 7)

When Can We Interpret β_1 Causally?

$\hat{\beta}_1$ always measures a **statistical association**. For it to be **causal**, we need:

The error term u_i must be **unrelated** to X_i in a specific sense:

$$\mathbb{E}[u \mid X] = 0$$

This is the **zero conditional mean** assumption — the single most important condition in this course.

When does it fail?

- **Omitted variables:** a factor in u is correlated with X
- **Reverse causality:** Y also affects X
- **Selection bias:** the sample is not representative conditional on X

We formalize this as Assumption SLR.4 later in this lecture.

Roadmap for This Lecture

1. The Simple Regression Model ✓
2. **Deriving the ordinary least squares (OLS) estimator**
3. **Assumptions for unbiasedness**
4. **Variance, efficiency, and the Gauss–Markov theorem**
5. **Goodness of fit:** R^2
6. **Preview of inference**

Reading: Wooldridge (2019, Chapters 1–2)

Part II
Deriving the OLS Estimator

Wooldridge, Chapter 2.2

The Idea: Best-Fitting Line

Goal: find the line $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ that "best fits" the data.

What does "best fit" mean?

- We want the line that makes the **residuals** $\hat{u}_i = Y_i - \hat{Y}_i$ as small as possible
- We cannot just minimize $\sum \hat{u}_i$ — positive and negative residuals cancel
- Instead, minimize the **sum of squared residuals**

Ordinary least squares (OLS): choose $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

Why squared? Penalizes large deviations; yields clean, closed-form solutions.

The OLS Objective Function

Define the sum of squared residuals as a function of $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$S(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

OLS problem:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \ S(\hat{\beta}_0, \hat{\beta}_1)$$

Strategy:

1. Take partial derivatives with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$
2. Set them equal to zero (first-order conditions)
3. Solve the resulting system of two equations in two unknowns

S is a convex quadratic in $(\hat{\beta}_0, \hat{\beta}_1)$, so the minimum is unique (if it exists).

First-Order Conditions

Taking partial derivatives and setting them to zero:

First-order condition (FOC) w.r.t. $\hat{\beta}_0$:

$$\frac{\partial S}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

FOC w.r.t. $\hat{\beta}_1$:

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n}X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

Dividing by -2 and rearranging gives the **normal equations**.

The Normal Equations

Rearranging the first-order conditions:

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

$$\sum_{i=1}^{n} X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

Interpretation:

- **First equation:** residuals sum to zero: $\sum_{i=1}^{n} \hat{u}_i = 0$
- **Second equation:** residuals are uncorrelated with X: $\sum_{i=1}^{n} X_i \hat{u}_i = 0$

These are the **sample analogs** of $\mathbb{E}[u] = 0$ and $\mathbb{E}[Xu] = 0$ — the foundation of the **method of moments**.

The OLS Estimators

Solving the normal equations yields closed-form solutions:

OLS estimator of the slope:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

OLS estimator of the intercept:

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

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ are the sample means. The OLS line always passes through the point $(\bar{X}, \bar{Y})$.

Intuition for $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\text{sample covariance of } X \text{ and } Y}{\text{sample variance of } X}$$

Reading the formula:

- **Numerator:** how X and Y move together (co-movement)
- **Denominator:** how much X varies (spread of X)
- $\hat{\beta}_1$ **scales** the co-movement by the variation in X

Connection to correlation:

$$\hat{\beta}_1 = \widehat{\text{Corr}}(X, Y) \cdot \frac{\hat{\sigma}_Y}{\hat{\sigma}_X}$$

If X and Y are positively correlated, $\hat{\beta}_1 > 0$. The stronger the correlation and the larger $\hat{\sigma}_Y / \hat{\sigma}_X$, the larger $|\hat{\beta}_1|$.

Fitted Values and Residuals

Fitted Values and Residuals

Fitted values: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ (predicted Y for observation i)

Residuals: $\hat{u}_i = Y_i - \hat{Y}_i$ (prediction error for observation i)

Key relationship:

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

Every observation decomposes into a **fitted part** (explained by the model) and a **residual** (unexplained).

Important distinction:

- **Errors** u_i: unobserved, theoretical
- **Residuals** $\hat{u}_i$: observed, computed from the sample

Algebraic Properties of OLS

These properties hold **by construction** — they follow from the normal equations, not from any assumptions about the data:

1. $\sum_{i=1}^{n} \hat{u}_i = 0$ (residuals sum to zero)
2. $\sum_{i=1}^{n} X_i \hat{u}_i = 0$ (residuals uncorrelated with X)
3. $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$ (line passes through $(\bar{X}, \bar{Y})$)
4. $\bar{\hat{Y}} = \bar{Y}$ (mean of fitted values equals mean of Y)

Properties 1–2 are the first-order conditions. Properties 3–4 follow directly.

OLS Derivation: Summary

OLS minimizes $\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$ and yields:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

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

What we have so far: a mechanical procedure to fit a line to data.

What we still need:

- Under what conditions does $\hat{\beta}_1$ tell us something meaningful about β_1?
- Is $\hat{\beta}_1$ unbiased? How precise is it? How well does the line fit?

This is where assumptions come in.

Part III
Assumptions and Properties

Wooldridge, Chapter 2.5

Why Do We Need Assumptions?

OLS is a **purely mechanical** procedure:

- You can always compute $\hat{\beta}_0$ and $\hat{\beta}_1$ from data
- OLS does not care where the data came from or what the model means
- Without assumptions, $\hat{\beta}_1$ is just a number — not an estimate of anything

Assumptions serve two purposes:

1. **Interpretation:** connect $\hat{\beta}_1$ to the population parameter β_1
2. **Properties:** guarantee desirable behavior (unbiasedness, efficiency, valid inference)

The assumptions we state are *sufficient conditions*. When they fail, OLS may still work — but we can no longer guarantee it.

Assumption SLR.1: Linear in Parameters

SLR.1 (Linear in Parameters): The population model is

$$Y = \beta_0 + \beta_1 X + u$$

where β_0 and β_1 are unknown parameters and u is the error term.

What "linear in parameters" means:

- Y is a *linear function* of β_0 and β_1
- X can be transformed: $\ln(X)$, X^2, $1/X$ are all fine
- $Y = \beta_0 + \beta_1 X^2 + u$ is linear in parameters
- $Y = \beta_0 + \beta_1^2 X + u$ **is not** linear in parameters

"Linear regression" refers to linearity in β, not in X.

Assumption SLR.2: Random Sampling

SLR.2 (Random Sampling): We have a random sample $\{(X_i, Y_i) : i = 1, \ldots, n\}$ from the population model.

This means:

- Each (X_i, Y_i) is drawn independently from the same joint distribution
- The relationship $Y_i = \beta_0 + \beta_1 X_i + u_i$ holds for every observation
- The parameters β_0, β_1 are the *same* for all i

When might this fail?

- **Time series:** observations are typically dependent over time
- **Clustered data:** students in the same school are correlated
- **Self-selection:** only certain people respond to a survey

Assumption SLR.3: Sample Variation in X

SLR.3 (Sample Variation): The sample values $X_1, \ldots, X_n$ are not all equal:

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

Why do we need this?

- Recall: $\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$
- If all X_i are equal, the denominator is **zero** — $\hat{\beta}_1$ is undefined
- **Intuition:** you cannot learn the effect of X on Y if X never varies

This is a *sample* requirement. More variation in X leads to more precise estimates — we formalize this when deriving $\mathrm{Var}(\hat{\beta}_1)$.

Assumption SLR.4: Zero Conditional Mean

SLR.4 (Zero Conditional Mean): The error has zero mean conditional on X:

$$\mathbb{E}[u \mid X] = 0$$

What this says:

- For *any* value of X, the average of the unobserved factors is zero
- Knowing X gives **no information** about u
- The population regression function is: $\mathbb{E}[Y \mid X] = \beta_0 + \beta_1 X$

Implications:

- $\mathbb{E}[u] = 0$ (by LIE: $\mathbb{E}[u] = \mathbb{E}[\mathbb{E}[u \mid X]] = 0$)
- $\mathrm{Cov}(X, u) = 0$ (implied, but the converse is *false*)

This is the **single most important assumption** in this course. Nearly every endogeneity problem reduces to a violation of $\mathbb{E}[u \mid X] = 0$.

Proving Unbiasedness of OLS

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

Step 1: Substitute $Y_i = \beta_0 + \beta_1 X_i + u_i$ into $\hat{\beta}_1$:

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

Step 2: Conditional expectation given $\mathbf{X} = (X_1, \ldots, X_n)$:

$$\mathbb{E}[\hat{\beta}_1 \mid \mathbf{X}] = \beta_1 + \frac{\sum(X_i - \bar{X})\,\mathbb{E}[u_i \mid \mathbf{X}]}{\sum(X_i - \bar{X})^2}$$

Step 3: By SLR.2 + SLR.4: $\mathbb{E}[u_i \mid \mathbf{X}] = \mathbb{E}[u_i \mid X_i] = 0$:

$\mathbb{E}[\hat{\beta}_1 \mid \mathbf{X}] = \beta_1$. By LIE: $\mathbb{E}[\hat{\beta}_1] = \mathbb{E}[\mathbb{E}[\hat{\beta}_1 \mid \mathbf{X}]] = \beta_1$. OLS is **unbiased**.

What $\mathbb{E}[u \mid X] = 0$ Rules Out

When $\mathbb{E}[u \mid X] \neq 0$, OLS is **biased**: $\mathbb{E}[\hat{\beta}_1] \neq \beta_1$.

1. Omitted variable bias — a factor in u is correlated with X

- Education example: ability is in u and correlates with educ
- $\hat{\beta}_1$ captures education effect $+$ ability effect $\Rightarrow$ upward bias

2. Reverse causality — Y also affects X

- Police and crime: more crime $\rightarrow$ more police $\rightarrow$ positive $\mathrm{Cov}(X, u)$

3. Measurement error — X is measured with noise

- If we observe $X = X^* + \varepsilon$ instead of the true X^*, the noise correlates with observed X

Solutions: multiple regression (L3), IV (L8), panel data (L10).

Assumption SLR.5: Homoskedasticity

SLR.5 (Homoskedasticity): The error has constant variance conditional on X:

$$\mathrm{Var}(u \mid X) = \sigma^2$$

Intuition: the "spread" of u around zero does not depend on X.

When might this fail?

- Wage variance often *increases* with education level
- Consumption variance often *increases* with income

When $\mathrm{Var}(u \mid X)$ varies with X, we have **heteroskedasticity** (Lecture 6).

SLR.5 is *not* needed for unbiasedness — only for the standard variance formula and the Gauss–Markov result.

Variance of $\hat{\beta}_1$

Under SLR.1–SLR.5, the conditional variance of $\hat{\beta}_1$ is:

$$\mathrm{Var}(\hat{\beta}_1 \mid \mathbf{X}) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sigma^2}{\mathsf{SST}_x}$$

where $\mathsf{SST}_x = \sum_{i=1}^{n}(X_i - \bar{X})^2$ is the total variation in X.

What makes $\hat{\beta}_1$ more precise?

1. **Smaller σ^2:** less noise in $u \Rightarrow$ sharper estimates
2. **Larger SST$_x$:** more variation in $X \Rightarrow$ more information about the slope
3. **Larger n:** more data $\Rightarrow$ SST$_x$ grows $\Rightarrow$ variance shrinks

Research design implication: choose X values with high spread when possible.

Estimating σ^2: The Standard Error of the Regression

The error variance $\sigma^2 = \mathrm{Var}(u \mid X)$ is unknown. We estimate it using residuals:

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

Why $n-2$?

- We lose 2 **degrees of freedom** from estimating β_0 and β_1
- The correction ensures unbiasedness: $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$

The **standard error of the regression** (SER) is:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

It measures the typical size of the OLS residuals — the "spread" of Y around the regression line, in the units of Y.

Standard Errors of the OLS Estimators

Replacing σ^2 with $\hat{\sigma}^2$ gives the estimated standard errors:

Slope: $$\text{se}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\text{SST}_x}}$$

Intercept: $$\text{se}(\hat{\beta}_0) = \hat{\sigma} \cdot \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n \cdot \text{SST}_x}}$$

Standard errors measure the precision of our estimates. They are used for:

- **Confidence intervals:** $\hat{\beta}_1 \pm t_{n-2,\,\alpha/2} \cdot \text{se}(\hat{\beta}_1)$
- **Hypothesis tests:** $t = \frac{\hat{\beta}_1 - \beta_{1,0}}{\text{se}(\hat{\beta}_1)}$

Full inference is covered in Lectures 3–4. For now, focus on what drives precision.

The Gauss–Markov Theorem

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).

What "BLUE" means:

- **Best:** lowest variance among all estimators that are…
- **Linear:** …linear functions of $Y_1, \ldots, Y_n$, and…
- **Unbiased:** …unbiased for β_j

Intuition: among all unbiased estimators that are linear in Y, OLS has the highest precision.

Caveat: Gauss–Markov does *not* say OLS is the best of all estimators — only the best *linear unbiased* one. Biased or nonlinear estimators can have lower mean squared error.

Assumption SLR.6: Normality of Errors

SLR.6 (Normality): $u \mid X \sim N(0, \sigma^2)$
Combined with SLR.1–SLR.5, this gives the **classical linear model** (CLM) assumptions.

What normality buys us:

- OLS estimators are *exactly* Normal: $\hat{\beta}_1 \mid \mathbf{X} \sim N\big(\beta_1,\ \sigma^2/\text{SST}_x\big)$
- The t-statistic follows an *exact* t-distribution (not just approximately)
- Enables exact finite-sample inference (no need for large n)

Without normality:

- OLS is still BLUE (Gauss–Markov only needs SLR.1–SLR.5)
- Inference is valid *asymptotically* via CLT (Lecture 5)
- For large n, the normality assumption matters very little

Summary: The SLR Assumptions

SLR.1: $Y = \beta_0 + \beta_1 X + u$ (linear in parameters)
SLR.2: $\{(X_i, Y_i)\}_{i=1}^{n}$ is a random sample (random sampling)
SLR.3: $\sum(X_i - \bar{X})^2 > 0$ (variation in X)
SLR.4: $\mathbb{E}[u \mid X] = 0$ (zero conditional mean)
SLR.5: $\mathrm{Var}(u \mid X) = \sigma^2$ (homoskedasticity)

What you get	Assumptions needed
OLS computable	SLR.1–SLR.3
Unbiasedness	SLR.1–SLR.4
Standard variance formula	SLR.1–SLR.5
BLUE (Gauss–Markov)	SLR.1–SLR.5
Exact t and F tests	SLR.1–SLR.5 $+$ SLR.6

Part V
Measuring Fit

Wooldridge, Chapter 2.3

Decomposing Total Variation in Y

Total variation: how much does Y vary around its mean?

$$Y_i - \bar{Y} = \underbrace{(\hat{Y}_i - \bar{Y})}_{\text{explained}} + \underbrace{\hat{u}_i}_{\text{residual}}$$

Squaring and summing (the cross-term vanishes since $\sum \hat{u}_i(\hat{Y}_i - \bar{Y}) = 0$):

$$\underbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}_{\text{TSS}} = \underbrace{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}_{\text{ESS}} + \underbrace{\sum_{i=1}^{n}\hat{u}_i^2}_{\text{RSS}}$$

- **TSS** — total sum of squares: total variation in Y
- **ESS** — explained sum of squares: variation explained by the model
- **RSS** — residual sum of squares: unexplained variation

R^2: The Coefficient of Determination

R^2

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

The **fraction of the total variation in Y explained by the regression**.

Properties:

- $0 \leq R^2 \leq 1$ (in simple regression with an intercept)
- $R^2 = 0$: the model explains *none* of the variation ($\hat{\beta}_1 = 0$)
- $R^2 = 1$: the model explains *all* of the variation (all points on the line)
- In simple regression: $R^2 = \left[\widehat{\text{Corr}}(Y, X)\right]^2$

What R^2 Does and Does Not Tell You

R^2 is useful for:

- Summarizing how well the model fits the data
- Comparing nested models (same dependent variable)

R^2 does *not* tell you:

- Whether $\hat{\beta}_1$ is causal
- Whether the model is correctly specified
- Whether the right variables are included
- That a "high" R^2 means the model is "good"

A low R^2 does *not* mean the regression is useless. If $\hat{\beta}_1$ is unbiased and precisely estimated, the effect of X on Y can be well-identified even when many other factors also affect Y.

Example: R^2 in the Wage Equation

Suppose we estimate:

$$\widehat{\ln(\text{wage})} = 0.584 + 0.083 \cdot \text{educ}, \qquad R^2 = 0.186$$

Interpretation of $R^2 = 0.186$:

- Education explains about 19% of the variation in log wages
- The other 81% is due to experience, ability, occupation, region, etc.

Is this "bad"? Not necessarily.

- In cross-sectional microdata, R^2 of 0.1–0.3 is common
- Humans differ in *many* ways — no single variable explains most of the variation
- $\hat{\beta}_1 = 0.083$ ($\approx 8.3\%$ return per year) is economically significant

Illustrative; based on typical Current Population Survey wage regressions. See Wooldridge (2019, Ch. 2).

Preview: Inference for β_1

With the SLR assumptions, we can **test hypotheses** and build **confidence intervals**:

t**-test** for $H_0: \beta_1 = 0$ ("X has no effect on Y"):

$$t = \frac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)} \sim t_{n-2} \quad \text{under } H_0 \text{ and SLR.1–SLR.6}$$

95% confidence interval for β_1:

$$\hat{\beta}_1 \pm t_{n-2,\,0.025} \cdot \text{se}(\hat{\beta}_1)$$

Full development in Lectures 3–4. Standard errors are the bridge from estimation to inference.

Next Time: Multiple Regression

In Lecture 3, we extend to multiple explanatory variables:

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

Why we need multiple regression:

- Control for confounding variables (reduce omitted variable bias)
- Isolate the effect of one variable, holding others constant
- Improve model fit and prediction

What we will cover: OLS in matrix form, interpreting coefficients, the Frisch–Waugh–Lovell theorem, R^2 and adjusted R^2.

Reading: Wooldridge (2019, Chapter 3)

Key Takeaways

1. The simple regression model $Y = \beta_0 + \beta_1 X + u$ relates an outcome to one explanatory variable
2. **OLS** minimizes $\sum \hat{u}_i^2$ and yields $\hat{\beta}_1 = \frac{\text{sample cov. of } X,Y}{\text{sample var. of } X}$
3. Under SLR.1–SLR.4, OLS is **unbiased**: $\mathbb{E}[\hat{\beta}_1] = \beta_1$
4. The **zero conditional mean** assumption $\mathbb{E}[u \mid X] = 0$ is crucial for causal interpretation
5. Under SLR.1–SLR.5, OLS is **BLUE** (Gauss–Markov theorem)
6. R^2 measures fit but says nothing about causality

References I

Mincer, J. (1974). *Schooling, Experience, and Earnings*. Columbia University Press, New York.

Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach*. Cengage Learning, Boston, MA, 7th edition.