Applied Regression

Francisco Villamil

Applied Quantitative Methods II

MA in Social Sciences, Spring 2026

Today's goals

- Review regression as modeling conditional expectations

Today's goals

- Review regression as modeling conditional expectations
- Understand OLS properties: assumptions, bias, standard errors

Today's goals

- Review regression as modeling conditional expectations
- Understand OLS properties: assumptions, bias, standard errors
- Understand multiple regression and control variables

Today's goals

- Review regression as modeling conditional expectations
- Understand OLS properties: assumptions, bias, standard errors
- Understand multiple regression and control variables
- Learn how to model conditional relationships (interactions)

Today's goals

- Review regression as modeling conditional expectations
- Understand OLS properties: assumptions, bias, standard errors
- Understand multiple regression and control variables
- Learn how to model conditional relationships (interactions)
- Present results effectively with `modelsummary`

Roadmap

Regression Review

OLS Properties

Multiple Regression

Interaction Effects

Presenting Results

What question does regression answer?

- "What is the average value of Y for different values of X?"

What question does regression answer?

- "What is the average value of Y for different values of X?"
- This is the **conditional expectation function** (CEF)

What question does regression answer?

- "What is the average value of Y for different values of X?"
- This is the **conditional expectation function** (CEF)
- Written as: $E[Y|X]$

What question does regression answer?

- "What is the average value of Y for different values of X?"
- This is the **conditional expectation function** (CEF)
- Written as: $E[Y|X]$
- Regression approximates this function

What does $E[\text{Income} \mid \text{Education}]$ look like?

Is it linear? Why or why not?

The regression model

The most common tool in social science:

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

- Y: outcome we want to explain

The regression model

The most common tool in social science:

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

- Y: outcome we want to explain
- X: explanatory variable(s)

The regression model

The most common tool in social science:

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

- Y: outcome we want to explain
- X: explanatory variable(s)
- β: coefficients (what we estimate)

The regression model

The most common tool in social science:

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

- Y: outcome we want to explain
- X: explanatory variable(s)
- β: coefficients (what we estimate)
- ε: error term (what we can't explain)

The regression model in matrix form

With n observations and k variables:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

The OLS estimator:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

The regression model in matrix form

With n observations and k variables:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

The OLS estimator:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Don't memorize – just know this is what `lm()` computes for you

Linear regression as approximation

- The true CEF might be complicated

Linear regression as approximation

- The true CEF might be complicated
- Linear regression fits the **best linear approximation**

Linear regression as approximation

- The true CEF might be complicated
- Linear regression fits the **best linear approximation**
- Even if the true relationship is non-linear

Linear regression as approximation

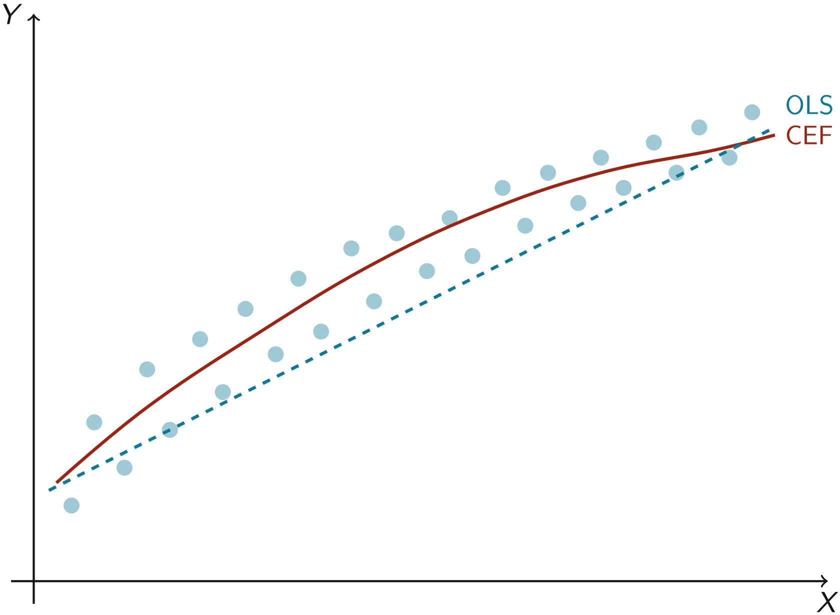
- The true CEF might be complicated
- Linear regression fits the **best linear approximation**
- Even if the true relationship is non-linear
- The linear fit is still the best predictor among linear functions

Linear regression as approximation

- The true CEF might be complicated
- Linear regression fits the **best linear approximation**

- Even if the true relationship is non-linear
- The linear fit is still the best predictor among linear functions

- Why linear? Simple, interpretable, often good enough


Y
OLS
CEF
X

Interpreting the slope coefficient

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

- β_1 represents:

Interpreting the slope coefficient

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

- β_1 represents:
  - $\rightarrow$ The difference in average Y

Interpreting the slope coefficient

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

- β_1 represents:
  - $\rightarrow$ The difference in average Y
  - $\rightarrow$ Between groups that differ by 1 unit in X

Interpreting the slope coefficient

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

- β_1 represents:
  - $\rightarrow$ The difference in average Y
  - $\rightarrow$ Between groups that differ by 1 unit in X

- This is a **comparison**, not necessarily a causal effect

Two framings of β_1

- **Predictive framing**:
  - $\rightarrow$ "Groups that differ by 1 in X differ by β_1 in Y, on average"
  - $\rightarrow$ A comparison across units

- **Counterfactual framing**:
  - $\rightarrow$ "If we changed X by 1, Y would change by β_1"
  - $\rightarrow$ A statement about what would happen

- Same number, very different claims
- The counterfactual framing requires **causal assumptions**

Which framing — predictive or counterfactual —

does a randomized experiment give you?

Descriptive vs. Causal interpretation

- **Descriptive**: How do units with different X values compare?
  - $\rightarrow$ "People with more education earn more, on average"
- **Causal**: What happens if we change X for a given unit?
  - $\rightarrow$ "If we give someone more education, they will earn more"
- Same coefficient, very different claims!

Running a regression in R

- The basic function: `lm(y ~ x, data = df)`

Running a regression in R

- The basic function: `lm(y ~ x, data = df)`
- Getting tidy output:

Running a regression in R

- The basic function: `lm(y ~ x, data = df)`
- Getting tidy output:
  - → `broom::tidy(model)` — coefficients as a data frame

Running a regression in R

- The basic function: `lm(y ~ x, data = df)`
- Getting tidy output:
  - → `broom::tidy(model)` — coefficients as a data frame
  - → `broom::glance(model)` — model-level statistics (R^2, etc.)

Running a regression in R

- The basic function: `lm(y ~ x, data = df)`
- Getting tidy output:
  - → `broom::tidy(model)` — coefficients as a data frame
  - → `broom::glance(model)` — model-level statistics (R^2, etc.)
- These are much easier to work with than `summary()`

Roadmap

OLS assumptions

For OLS to work well, we need:

1. **Linearity**: $Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

OLS assumptions

For OLS to work well, we need:

1. **Linearity**: $Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
2. **Random sampling**: observations are i.i.d.

OLS assumptions

For OLS to work well, we need:

1. **Linearity**: $Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
2. **Random sampling**: observations are i.i.d.
3. **No perfect multicollinearity**: $\mathbf{X}'\mathbf{X}$ is invertible

OLS assumptions

For OLS to work well, we need:

1. **Linearity**: $Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
2. **Random sampling**: observations are i.i.d.
3. **No perfect multicollinearity**: $\mathbf{X}'\mathbf{X}$ is invertible
4. **Zero conditional mean**: $E[\varepsilon|\mathbf{X}] = 0$

OLS assumptions

For OLS to work well, we need:

1. **Linearity**: $Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
2. **Random sampling**: observations are i.i.d.
3. **No perfect multicollinearity**: $\mathbf{X}'\mathbf{X}$ is invertible
4. **Zero conditional mean**: $E[\varepsilon|\mathbf{X}] = 0$
5. **Homoskedasticity**: $\text{Var}(\varepsilon|\mathbf{X}) = \sigma^2$

OLS assumptions

For OLS to work well, we need:

1. **Linearity**: $Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
2. **Random sampling**: observations are i.i.d.
3. **No perfect multicollinearity**: $\mathbf{X}'\mathbf{X}$ is invertible
4. **Zero conditional mean**: $E[\varepsilon|\mathbf{X}] = 0$
5. **Homoskedasticity**: $\mathrm{Var}(\varepsilon|\mathbf{X}) = \sigma^2$

OLS assumptions

For OLS to work well, we need:

1. **Linearity**: $Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
2. **Random sampling**: observations are i.i.d.
3. **No perfect multicollinearity**: $\mathbf{X}'\mathbf{X}$ is invertible
4. **Zero conditional mean**: $E[\varepsilon|\mathbf{X}] = 0$
5. **Homoskedasticity**: $\text{Var}(\varepsilon|\mathbf{X}) = \sigma^2$

A1–A4 are needed for unbiasedness; A5 for efficient SEs

OLS is unbiased (under A1–A4)

If assumptions A1–A4 hold:

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

- On average, OLS gives us the right answer

OLS is unbiased (under A1–A4)

If assumptions A1–A4 hold:

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

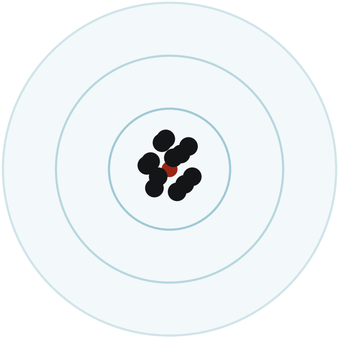
- On average, OLS gives us the right answer
- Any single estimate might be off, but there is no systematic error

OLS is unbiased (under A1–A4)

If assumptions A1–A4 hold:

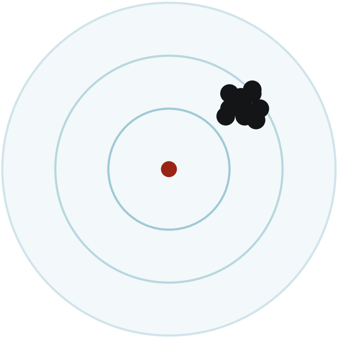
$$E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$$

- On average, OLS gives us the right answer
- Any single estimate might be off, but there is no systematic error
- *Think of it like an unbiased dart thrower:*
  centered on the bullseye, but with some scatter



Unbiased

centered on target



Biased

systematically off

Standard errors and uncertainty

OLS gives us $\hat{\beta}$, but how precise is it?

$$SE(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{SST_j(1 - R_j^2)}}$$

- $\hat{\sigma}^2$: error variance (more noise $\rightarrow$ larger SE)

Standard errors and uncertainty

OLS gives us $\hat{\beta}$, but how precise is it?

$$SE(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{SST_j(1 - R_j^2)}}$$

- $\hat{\sigma}^2$: error variance (more noise $\rightarrow$ larger SE)
- SST_j: variation in X_j (more variation $\rightarrow$ smaller SE)

Standard errors and uncertainty

OLS gives us $\hat{\beta}$, but how precise is it?

$$SE(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{SST_j(1 - R_j^2)}}$$

- $\hat{\sigma}^2$: error variance (more noise $\rightarrow$ larger SE)
- SST_j: variation in X_j (more variation $\rightarrow$ smaller SE)
- R_j^2: correlation of X_j with other predictors (multicollinearity $\rightarrow$ larger SE)

Standard errors and uncertainty

OLS gives us $\hat{\beta}$, but how precise is it?

$$SE(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{SST_j(1 - R_j^2)}}$$

- $\hat{\sigma}^2$: error variance (more noise $\rightarrow$ larger SE)
- SST_j: variation in X_j (more variation $\rightarrow$ smaller SE)
- R_j^2: correlation of X_j with other predictors (multicollinearity $\rightarrow$ larger SE)

- The SE tells us how much $\hat{\beta}$ would vary across samples

Heteroskedasticity

- Assumption A5 says the error variance is constant

Heteroskedasticity

- Assumption A5 says the error variance is constant
- In practice, it almost never is

Heteroskedasticity

- Assumption A5 says the error variance is constant
- In practice, it almost never is
- **Heteroskedasticity**: $\text{Var}(\varepsilon|X)$ changes with X

Heteroskedasticity

- Assumption A5 says the error variance is constant
- In practice, it almost never is
- **Heteroskedasticity**: $\mathrm{Var}(\varepsilon|X)$ changes with X
- Example: income variation is larger for people with more education

Heteroskedasticity

- Assumption A5 says the error variance is constant
- In practice, it almost never is
- **Heteroskedasticity**: $\text{Var}(\varepsilon|X)$ changes with X
- Example: income variation is larger for people with more education
- The estimates $\hat{\beta}$ are still unbiased!

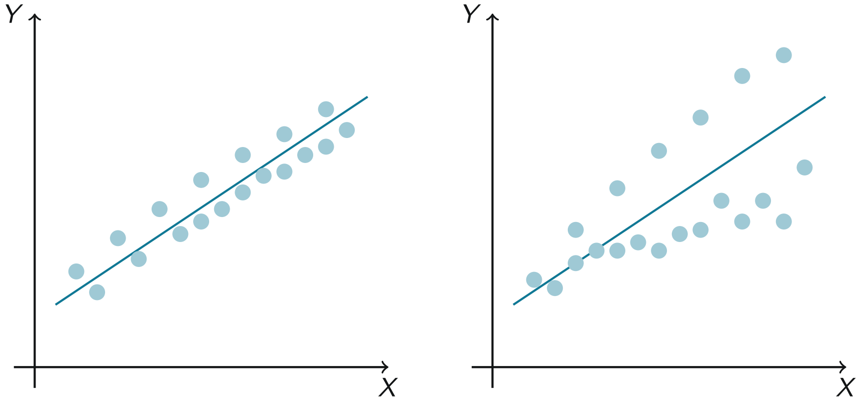
Heteroskedasticity

- Assumption A5 says the error variance is constant
- In practice, it almost never is

- **Heteroskedasticity**: $\text{Var}(\varepsilon|X)$ changes with X

- Example: income variation is larger for people with more education

- The estimates $\hat{\beta}$ are still unbiased!
- But the standard errors are wrong



Homoskedastic
constant spread

Heteroskedastic
spread increases with X

Solution: robust standard errors

- Robust ("sandwich") SEs are valid even with heteroskedasticity

Solution: robust standard errors

- Robust ("sandwich") SEs are valid even with heteroskedasticity
- Also called HC (heteroskedasticity-consistent) standard errors

Solution: robust standard errors

- Robust ("sandwich") SEs are valid even with heteroskedasticity
- Also called HC (heteroskedasticity-consistent) standard errors
- In R with `modelsummary`:

Solution: robust standard errors

- Robust ("sandwich") SEs are valid even with heteroskedasticity
- Also called HC (heteroskedasticity-consistent) standard errors
- In R with `modelsummary`:
  - → `modelsummary(model, vcov = "robust")`

Solution: robust standard errors

- Robust ("sandwich") SEs are valid even with heteroskedasticity
- Also called HC (heteroskedasticity-consistent) standard errors
- In R with `modelsummary`:

  → `modelsummary(model, vcov = "robust")`

- Or using `lmtest` and `sandwich`:

Solution: robust standard errors

- Robust ("sandwich") SEs are valid even with heteroskedasticity
- Also called HC (heteroskedasticity-consistent) standard errors
- In R with `modelsummary`:
  - → `modelsummary(model, vcov = "robust")`
- Or using `lmtest` and `sandwich`:
  - → `coeftest(model, vcov = vcovHC(model))`

Solution: robust standard errors

- Robust ("sandwich") SEs are valid even with heteroskedasticity
- Also called HC (heteroskedasticity-consistent) standard errors
- In R with `modelsummary`:
  $\rightarrow$ `modelsummary(model, vcov = "robust")`
- Or using `lmtest` and `sandwich`:
  $\rightarrow$ `coeftest(model, vcov = vcovHC(model))`
- **Practical rule**: always use robust SEs

Solution: robust standard errors

- Robust ("sandwich") SEs are valid even with heteroskedasticity
- Also called HC (heteroskedasticity-consistent) standard errors
- In R with `modelsummary`:

  → `modelsummary(model, vcov = "robust")`
- Or using `lmtest` and `sandwich`:

  → `coeftest(model, vcov = vcovHC(model))`
- **Practical rule**: always use robust SEs
- There is no real cost when errors are homoskedastic

When robust SEs are not enough: clustering

- Robust SEs fix heteroskedasticity

When robust SEs are not enough: clustering

- Robust SEs fix heteroskedasticity
- But what if observations are **not independent**?

When robust SEs are not enough: clustering

- Robust SEs fix heteroskedasticity
- But what if observations are **not independent**?
- Examples of clustered data:

When robust SEs are not enough: clustering

- Robust SEs fix heteroskedasticity
- But what if observations are **not independent**?
- Examples of clustered data:
  → Students within schools

When robust SEs are not enough: clustering

- Robust SEs fix heteroskedasticity
- But what if observations are **not independent**?
- Examples of clustered data:
  - → Students within schools
  - → Citizens within countries

When robust SEs are not enough: clustering

- Robust SEs fix heteroskedasticity
- But what if observations are **not independent**?
- Examples of clustered data:
  - $\rightarrow$ Students within schools
  - $\rightarrow$ Citizens within countries
  - $\rightarrow$ Repeated observations of the same unit over time

When robust SEs are not enough: clustering

- Robust SEs fix heteroskedasticity
- But what if observations are **not independent**?
- Examples of clustered data:
  - → Students within schools
  - → Citizens within countries
  - → Repeated observations of the same unit over time
- Within a cluster, errors are correlated

When robust SEs are not enough: clustering

- Robust SEs fix heteroskedasticity
- But what if observations are **not independent**?
- Examples of clustered data:
  - → Students within schools
  - → Citizens within countries
  - → Repeated observations of the same unit over time
- Within a cluster, errors are correlated
- Standard SEs (even robust) are **too small**

When robust SEs are not enough: clustering

- Robust SEs fix heteroskedasticity
- But what if observations are **not independent**?
- Examples of clustered data:
  - → Students within schools
  - → Citizens within countries
  - → Repeated observations of the same unit over time
- Within a cluster, errors are correlated
- Standard SEs (even robust) are **too small**
- This means too many “significant” results

Clustered standard errors in practice

- **Rule of thumb**: cluster at the level of treatment assignment

Clustered standard errors in practice

- **Rule of thumb**: cluster at the level of treatment assignment
  - $\rightarrow$ Policy varies by country? Cluster by country

Clustered standard errors in practice

- **Rule of thumb**: cluster at the level of treatment assignment
  - → Policy varies by country? Cluster by country
  - → Treatment assigned by school? Cluster by school

Clustered standard errors in practice

- **Rule of thumb**: cluster at the level of treatment assignment
  - $\rightarrow$ Policy varies by country? Cluster by country
  - $\rightarrow$ Treatment assigned by school? Cluster by school
- In R with `modelsummary`:

Clustered standard errors in practice

- **Rule of thumb**: cluster at the level of treatment assignment
  - → Policy varies by country? Cluster by country
  - → Treatment assigned by school? Cluster by school
- In R with `modelsummary`:
  - → `modelsummary(model, vcov = ~country)`

Clustered standard errors in practice

- **Rule of thumb**: cluster at the level of treatment assignment
  - $\rightarrow$ Policy varies by country? Cluster by country
  - $\rightarrow$ Treatment assigned by school? Cluster by school
- In R with `modelsummary`:
  - $\rightarrow$ `modelsummary(model, vcov = ~country)`
- Clustered SEs $\geq$ robust SEs (always at least as large)

Clustered standard errors in practice

- **Rule of thumb**: cluster at the level of treatment assignment
  - $\rightarrow$ Policy varies by country? Cluster by country
  - $\rightarrow$ Treatment assigned by school? Cluster by school
- In R with `modelsummary`:
  - $\rightarrow$ `modelsummary(model, vcov = ~country)`
- Clustered SEs $\geq$ robust SEs (always at least as large)
- With few clusters ($<$ 30–50), even clustered SEs can be unreliable

Roadmap

Regression Review

OLS Properties

Multiple Regression

Interaction Effects

Presenting Results

Adding predictors

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

- β_1 now represents:

Adding predictors

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

- β_1 now represents:
  - $\rightarrow$ The difference in average Y

Adding predictors

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

- β_1 now represents:
  - $\rightarrow$ The difference in average Y
  - $\rightarrow$ Between groups that differ by 1 in X_1

Adding predictors

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

- β_1 now represents:
  - → The difference in average Y
  - → Between groups that differ by 1 in X_1
  - → **Holding X_2 constant**

Adding predictors

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

- β_1 now represents:
  - $\rightarrow$ The difference in average Y
  - $\rightarrow$ Between groups that differ by 1 in X_1
  - $\rightarrow$ **Holding X_2 constant**

- This is the "controlled" effect of X_1

How controlling works

- OLS with multiple variables "partials out" the controls

How controlling works

- OLS with multiple variables “partials out” the controls
- Technically: we look at variation in X_1 that is unrelated to X_2

How controlling works

- OLS with multiple variables "partials out" the controls
- Technically: we look at variation in X_1 that is unrelated to X_2
- This isolates the unique contribution of X_1

Omitted variable bias

If we omit a relevant variable X_2, the short regression gives:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \cdot \tilde{\delta}$$

- $\hat{\beta}_1$: the “true” coefficient from the long regression

Omitted variable bias

If we omit a relevant variable X_2, the short regression gives:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \cdot \tilde{\delta}$$

- $\hat{\beta}_1$: the "true" coefficient from the long regression
- $\hat{\beta}_2$: the effect of the omitted variable on Y

Omitted variable bias

If we omit a relevant variable X_2, the short regression gives:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \cdot \tilde{\delta}$$

- $\hat{\beta}_1$: the "true" coefficient from the long regression
- $\hat{\beta}_2$: the effect of the omitted variable on Y
- $\tilde{\delta}$: the relationship between X_2 and X_1

Omitted variable bias

If we omit a relevant variable X_2, the short regression gives:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \cdot \tilde{\delta}$$

- $\hat{\beta}_1$: the "true" coefficient from the long regression
- $\hat{\beta}_2$: the effect of the omitted variable on Y
- $\tilde{\delta}$: the relationship between X_2 and X_1
  - $\rightarrow$ (coefficient from regressing X_2 on X_1)

Omitted variable bias

If we omit a relevant variable X_2, the short regression gives:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \cdot \tilde{\delta}$$

- $\hat{\beta}_1$: the "true" coefficient from the long regression
- $\hat{\beta}_2$: the effect of the omitted variable on Y
- $\tilde{\delta}$: the relationship between X_2 and X_1
  - $\rightarrow$ (coefficient from regressing X_2 on X_1)
- Bias $= \hat{\beta}_2 \cdot \tilde{\delta}$

Omitted variable bias

If we omit a relevant variable X_2, the short regression gives:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \cdot \tilde{\delta}$$

- $\hat{\beta}_1$: the "true" coefficient from the long regression
- $\hat{\beta}_2$: the effect of the omitted variable on Y
- $\tilde{\delta}$: the relationship between X_2 and X_1
  - $\rightarrow$ (coefficient from regressing X_2 on X_1)

- Bias $= \hat{\beta}_2 \cdot \tilde{\delta}$
- Zero only if $\hat{\beta}_2 = 0$ *or* $\tilde{\delta} = 0$

OVB in practice: education and income

	Short regression (omits ability)	Long regression (includes ability)
Education (β_1)	\$5,000	\$3,000
Ability (β_2)	—	\$5,000

- Auxiliary regression: $\tilde{\delta} = 0.4$ (ability on education)

OVB in practice: education and income

	Short regression (omits ability)	Long regression (includes ability)
Education (β_1)	\$5,000	\$3,000
Ability (β_2)	—	\$5,000

- Auxiliary regression: $\tilde{\delta} = 0.4$ (ability on education)
- Check: $\underbrace{\$3{,}000}_{\hat{\beta}_1} + \underbrace{\$5{,}000}_{\hat{\beta}_2} \times \underbrace{0.4}_{\tilde{\delta}} = \underbrace{\$5{,}000}_{\tilde{\beta}_1}$ ✓

OVB in practice: education and income

	Short regression (omits ability)	Long regression (includes ability)
Education (β_1)	\$5,000	\$3,000
Ability (β_2)	—	\$5,000

- Auxiliary regression: $\tilde{\delta} = 0.4$ (ability on education)
- Check: $\underbrace{\$3{,}000}_{\hat{\beta}_1} + \underbrace{\$5{,}000}_{\hat{\beta}_2} \times \underbrace{0.4}_{\tilde{\delta}} = \underbrace{\$5{,}000}_{\tilde{\beta}_1}$ ✓
- Bias $=$ \$2,000 — we overestimate by 67%!

OVB in practice: education and income

	Short regression (omits ability)	Long regression (includes ability)
Education (β_1)	\$5,000	\$3,000
Ability (β_2)	—	\$5,000

- Auxiliary regression: $\tilde{\delta} = 0.4$ (ability on education)
- Check: $\underbrace{\$3{,}000}_{\hat{\beta}_1} + \underbrace{\$5{,}000}_{\hat{\beta}_2} \times \underbrace{0.4}_{\tilde{\delta}} = \underbrace{\$5{,}000}_{\tilde{\beta}_1}$ ✓
- Bias $=$ \$2,000 — we overestimate by 67%!
- Because ability $\uparrow$ education *and* ability $\uparrow$ income

What makes a good control?

Good controls are variables that:

- Affect both the treatment and the outcome

What makes a good control?

Good controls are variables that:

- Affect both the treatment and the outcome
- Are determined **before** the treatment

What makes a good control?

Good controls are variables that:

- Affect both the treatment and the outcome
- Are determined **before** the treatment
- Are not affected by the treatment

What makes a good control?

Good controls are variables that:

- Affect both the treatment and the outcome
- Are determined **before** the treatment
- Are not affected by the treatment

What makes a good control?

Good controls are variables that:

- Affect both the treatment and the outcome
- Are determined **before** the treatment
- Are not affected by the treatment

Pre-treatment confounders are the key!

You study the effect of job training on wages.

Is *current job type* a good or bad control?

Why?

Bad controls: Post-treatment variables

- Never control for variables caused by the treatment

Bad controls: Post-treatment variables

- Never control for variables caused by the treatment
- Example: Studying effect of job training on wages

Bad controls: Post-treatment variables

- Never control for variables caused by the treatment
- Example: Studying effect of job training on wages
  → Don't control for job type (affected by training)

Bad controls: Post-treatment variables

- Never control for variables caused by the treatment
- Example: Studying effect of job training on wages
  - → Don't control for job type (affected by training)
  - → Do control for education (determined before training)

Bad controls: Post-treatment variables

- Never control for variables caused by the treatment
- Example: Studying effect of job training on wages
  - → Don't control for job type (affected by training)
  - → Do control for education (determined before training)
- Controlling for post-treatment variables can *introduce* bias

Bad controls: Colliders

- A **collider** is caused by both X and Y

Bad controls: Colliders

- A **collider** is caused by both X and Y
- Controlling for it creates a spurious association

Bad controls: Colliders

- A **collider** is caused by both X and Y
- Controlling for it creates a spurious association
- Example: NBA players

Bad controls: Colliders

- A **collider** is caused by both X and Y
- Controlling for it creates a spurious association
- Example: NBA players
  - $\rightarrow$ Height and skill both affect being in NBA

Bad controls: Colliders

- A **collider** is caused by both X and Y
- Controlling for it creates a spurious association
- Example: NBA players
  - $\rightarrow$ Height and skill both affect being in NBA
  - $\rightarrow$ Among NBA players, height and skill are negatively correlated

Bad controls: Colliders

- A **collider** is caused by both X and Y
- Controlling for it creates a spurious association
- Example: NBA players
  - $\rightarrow$ Height and skill both affect being in NBA
  - $\rightarrow$ Among NBA players, height and skill are negatively correlated
  - $\rightarrow$ But not in the general population!

Categorical predictors

- What if X is a category (region, party, gender)?

Categorical predictors

- What if X is a category (region, party, gender)?
- R automatically creates **dummy variables**

Categorical predictors

- What if X is a category (region, party, gender)?
- R automatically creates **dummy variables**
  → One indicator (0/1) for each category

Categorical predictors

- What if X is a category (region, party, gender)?
- R automatically creates **dummy variables**
  - $\rightarrow$ One indicator (0/1) for each category
  - $\rightarrow$ One category is the **reference** (omitted)

Categorical predictors

- What if X is a category (region, party, gender)?
- R automatically creates **dummy variables**
  - $\rightarrow$ One indicator (0/1) for each category
  - $\rightarrow$ One category is the **reference** (omitted)
- Coefficients represent the difference from the reference

Categorical predictors

- What if X is a category (region, party, gender)?
- R automatically creates **dummy variables**
  - $\rightarrow$ One indicator (0/1) for each category
  - $\rightarrow$ One category is the **reference** (omitted)
- Coefficients represent the difference from the reference
- Example: `lm(income ~ factor(region), data = df)`

Categorical predictors

- What if X is a category (region, party, gender)?
- R automatically creates **dummy variables**
  - $\rightarrow$ One indicator (0/1) for each category
  - $\rightarrow$ One category is the **reference** (omitted)
- Coefficients represent the difference from the reference
- Example: `lm(income ~ factor(region), data = df)`
  - $\rightarrow$ If reference is "North", the "South" coefficient means: average income in South minus average income in North

Roadmap

Regression Review

OLS Properties

Multiple Regression

Interaction Effects

Presenting Results

When effects depend on context

- Sometimes, the effect of X on Y depends on another variable Z

When effects depend on context

- Sometimes, the effect of X on Y depends on another variable Z
- Examples:

When effects depend on context

- Sometimes, the effect of X on Y depends on another variable Z
- Examples:
  - $\rightarrow$ Effect of education on income may differ by gender

When effects depend on context

- Sometimes, the effect of X on Y depends on another variable Z
- Examples:
  - → Effect of education on income may differ by gender
  - → Effect of campaign spending may differ by incumbency status

When effects depend on context

- Sometimes, the effect of X on Y depends on another variable Z
- Examples:
  - $\rightarrow$ Effect of education on income may differ by gender
  - $\rightarrow$ Effect of campaign spending may differ by incumbency status
  - $\rightarrow$ Effect of democracy on growth may depend on economic development

When effects depend on context

- Sometimes, the effect of X on Y depends on another variable Z
- Examples:
  - $\rightarrow$ Effect of education on income may differ by gender
  - $\rightarrow$ Effect of campaign spending may differ by incumbency status
  - $\rightarrow$ Effect of democracy on growth may depend on economic development
- We model this with **interaction terms**

The interaction model

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 (X \times Z) + \varepsilon$$

- β_1: effect of X when $Z = 0$

The interaction model

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 (X \times Z) + \varepsilon$$

- β_1: effect of X when $Z = 0$
- β_2: effect of Z when $X = 0$

The interaction model

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 (X \times Z) + \varepsilon$$

- β_1: effect of X when $Z = 0$
- β_2: effect of Z when $X = 0$
- β_3: how the effect of X changes as Z increases

The marginal effect of X

$$\frac{\partial Y}{\partial X} = \beta_1 + \beta_3 Z$$

- The effect of X is no longer a single number

The marginal effect of X

$$\frac{\partial Y}{\partial X} = \beta_1 + \beta_3 Z$$

- The effect of X is no longer a single number
- It's a **function** of Z

The marginal effect of X

$$\frac{\partial Y}{\partial X} = \beta_1 + \beta_3 Z$$

- The effect of X is no longer a single number
- It's a **function** of Z
- Need to report effects at meaningful values of Z

Continuous $\times$ categorical interactions

- When Z is categorical (e.g., gender, regime type)

Continuous × categorical interactions

- When Z is categorical (e.g., gender, regime type)
- The interaction gives a **different slope** for each group

Continuous $\times$ categorical interactions

- When Z is categorical (e.g., gender, regime type)
- The interaction gives a **different slope** for each group
- Example: `lm(income ~ education * gender, data = df)`

Continuous $\times$ categorical interactions

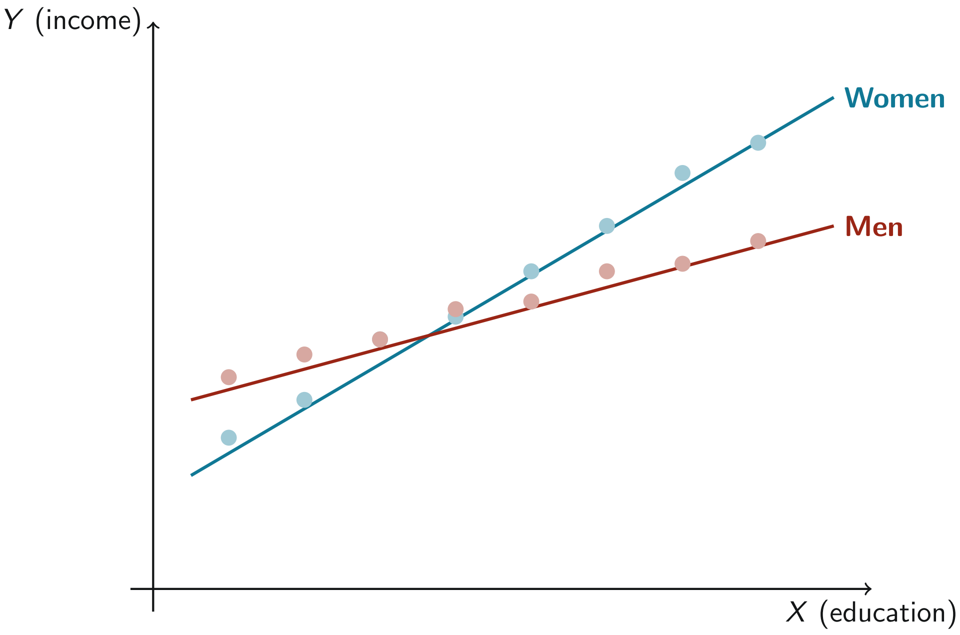
- When Z is categorical (e.g., gender, regime type)
- The interaction gives a **different slope** for each group
- Example: `lm(income ~ education * gender, data = df)`
  - $\rightarrow$ One slope for men, a different slope for women

Continuous × categorical interactions

- When Z is categorical (e.g., gender, regime type)
- The interaction gives a **different slope** for each group
- Example: `lm(income ~ education * gender, data = df)`
  - → One slope for men, a different slope for women
- Equivalent to fitting separate regressions by group

Continuous × categorical interactions

- When Z is categorical (e.g., gender, regime type)
- The interaction gives a **different slope** for each group
- Example: `lm(income ~ education * gender, data = df)`
  → One slope for men, a different slope for women
- Equivalent to fitting separate regressions by group
- But estimated jointly (shares the error variance)


Y (income)
Women
Men
X (education)

Continuous $\times$ continuous interactions

- When both X and Z are continuous

Continuous $\times$ continuous interactions

- When both X and Z are continuous
- The slope of X varies smoothly with Z (and vice versa)

Continuous $\times$ continuous interactions

- When both X and Z are continuous
- The slope of X varies smoothly with Z (and vice versa)

- Harder to interpret from coefficients alone

Continuous $\times$ continuous interactions

- When both X and Z are continuous
- The slope of X varies smoothly with Z (and vice versa)
- Harder to interpret from coefficients alone
- Best communicated through plots:

Continuous × continuous interactions

- When both X and Z are continuous
- The slope of X varies smoothly with Z (and vice versa)
- Harder to interpret from coefficients alone
- Best communicated through plots:
  - → Predicted values at different combinations of X and Z

Continuous $\times$ continuous interactions

- When both X and Z are continuous
- The slope of X varies smoothly with Z (and vice versa)
- Harder to interpret from coefficients alone
- Best communicated through plots:
  - $\rightarrow$ Predicted values at different combinations of X and Z
  - $\rightarrow$ Marginal effect of X across values of Z

Common mistakes with interactions

- **Mistake 1**: Interpreting β_1 as "the effect of X"

Common mistakes with interactions

- **Mistake 1**: Interpreting β_1 as "the effect of X"
  - $\rightarrow$ It's only the effect when $Z = 0$

Common mistakes with interactions

- **Mistake 1**: Interpreting β_1 as "the effect of X"
  - $\rightarrow$ It's only the effect when $Z = 0$
  - $\rightarrow$ May not even be meaningful!

Common mistakes with interactions

- **Mistake 1**: Interpreting β_1 as "the effect of X"
  - $\rightarrow$ It's only the effect when $Z = 0$
  - $\rightarrow$ May not even be meaningful!
- **Mistake 2**: Omitting constitutive terms

Common mistakes with interactions

- **Mistake 1**: Interpreting β_1 as "the effect of X"
  - $\rightarrow$ It's only the effect when $Z = 0$
  - $\rightarrow$ May not even be meaningful!
- **Mistake 2**: Omitting constitutive terms
  - $\rightarrow$ Always include X and Z separately, not just $X \times Z$

Common mistakes with interactions

- **Mistake 1**: Interpreting β_1 as "the effect of X"
  - $\rightarrow$ It's only the effect when $Z = 0$
  - $\rightarrow$ May not even be meaningful!
- **Mistake 2**: Omitting constitutive terms
  - $\rightarrow$ Always include X and Z separately, not just $X \times Z$
- **Mistake 3**: Not showing how the effect varies

Common mistakes with interactions

- **Mistake 1**: Interpreting β_1 as "the effect of X"
  - $\rightarrow$ It's only the effect when $Z = 0$
  - $\rightarrow$ May not even be meaningful!
- **Mistake 2**: Omitting constitutive terms
  - $\rightarrow$ Always include X and Z separately, not just $X \times Z$
- **Mistake 3**: Not showing how the effect varies
  - $\rightarrow$ Plot the marginal effect across values of Z

Visualizing interactions

- Tables of coefficients are hard to interpret

Visualizing interactions

- Tables of coefficients are hard to interpret
- Better approach:

Visualizing interactions

- Tables of coefficients are hard to interpret
- Better approach:
  - $\rightarrow$ Plot predicted values of Y for different combinations of X and Z

Visualizing interactions

- Tables of coefficients are hard to interpret
- Better approach:
  - $\rightarrow$ Plot predicted values of Y for different combinations of X and Z
  - $\rightarrow$ Plot the marginal effect of X across values of Z

Visualizing interactions

- Tables of coefficients are hard to interpret
- Better approach:
  - $\rightarrow$ Plot predicted values of Y for different combinations of X and Z
  - $\rightarrow$ Plot the marginal effect of X across values of Z
  - $\rightarrow$ Include confidence intervals

Visualizing interactions

- Tables of coefficients are hard to interpret
- Better approach:
  - $\rightarrow$ Plot predicted values of Y for different combinations of X and Z
  - $\rightarrow$ Plot the marginal effect of X across values of Z
  - $\rightarrow$ Include confidence intervals
- In R: `marginaleffects::plot_predictions()`

Roadmap

Why presentation matters

- A regression table is not the end of the analysis

Why presentation matters

- A regression table is not the end of the analysis
- Readers need to understand the **substance** of your findings

Why presentation matters

- A regression table is not the end of the analysis
- Readers need to understand the **substance** of your findings
- Good presentation:

Why presentation matters

- A regression table is not the end of the analysis
- Readers need to understand the **substance** of your findings
- Good presentation:
  - → Shows what the results **mean**, not just what they are

Why presentation matters

- A regression table is not the end of the analysis
- Readers need to understand the **substance** of your findings
- Good presentation:
  - → Shows what the results **mean**, not just what they are
  - → Communicates **uncertainty** honestly

Why presentation matters

- A regression table is not the end of the analysis
- Readers need to understand the **substance** of your findings
- Good presentation:
  - → Shows what the results **mean**, not just what they are
  - → Communicates **uncertainty** honestly
  - → Helps readers evaluate the **size** of effects

The `modelsummary` package

- Creates publication-quality tables from model objects

The `modelsummary` package

- Creates publication-quality tables from model objects
- Basic usage:

The `modelsummary` package

- Creates publication-quality tables from model objects
- Basic usage:
  - → `modelsummary(model)`

The `modelsummary` package

- Creates publication-quality tables from model objects
- Basic usage:
  - → `modelsummary(model)`
  - → `modelsummary(list(m1, m2, m3))`

The `modelsummary` package

- Creates publication-quality tables from model objects
- Basic usage:
  - → `modelsummary(model)`
  - → `modelsummary(list(m1, m2, m3))`
- Output formats: LaTeX, HTML, Word, markdown

The `modelsummary` package

- Creates publication-quality tables from model objects
- Basic usage:
  - → `modelsummary(model)`
  - → `modelsummary(list(m1, m2, m3))`
- Output formats: LaTeX, HTML, Word, markdown
- Highly customizable: statistics, labels, notes

Coefficient plots

- A visual alternative to tables

Coefficient plots

- A visual alternative to tables
- `modelsummary::modelplot(model)`

Coefficient plots

- A visual alternative to tables
- `modelsummary::modelplot(model)`
  → Each coefficient as a point with confidence interval

Coefficient plots

- A visual alternative to tables
- `modelsummary::modelplot(model)`
  - → Each coefficient as a point with confidence interval
  - → Easy to compare multiple models

Coefficient plots

- A visual alternative to tables
- `modelsummary::modelplot(model)`
  - → Each coefficient as a point with confidence interval
  - → Easy to compare multiple models
- Often more effective than tables for communicating results

Coefficient plots

- A visual alternative to tables
- `modelsummary::modelplot(model)`
  - → Each coefficient as a point with confidence interval
  - → Easy to compare multiple models
- Often more effective than tables for communicating results
- Readers immediately see which effects are large vs. small

Building sequential models

- Common strategy: show how results change as you add variables

Building sequential models

- Common strategy: show how results change as you add variables
- Step 1: Bivariate model (just X and Y)

Building sequential models

- Common strategy: show how results change as you add variables
- Step 1: Bivariate model (just X and Y)
- Step 2: Add control variables

Building sequential models

- Common strategy: show how results change as you add variables
- Step 1: Bivariate model (just X and Y)
- Step 2: Add control variables
- Step 3: Add interactions

Building sequential models

- Common strategy: show how results change as you add variables

- Step 1: Bivariate model (just X and Y)
- Step 2: Add control variables
- Step 3: Add interactions

- Present all three in one table:

Building sequential models

- Common strategy: show how results change as you add variables

- Step 1: Bivariate model (just X and Y)
- Step 2: Add control variables
- Step 3: Add interactions

- Present all three in one table:

  $\rightarrow$ `modelsummary(list(m1, m2, m3))`

Building sequential models

- Common strategy: show how results change as you add variables

- Step 1: Bivariate model (just X and Y)
- Step 2: Add control variables
- Step 3: Add interactions

- Present all three in one table:

  $\rightarrow$ `modelsummary(list(m1, m2, m3))`

- Shows robustness and what adding controls does to the estimate

Example workflow in R

```
m1 <- lm(y ~ x, data = df)
m2 <- lm(y ~ x + z1 + z2, data = df)
m3 <- lm(y ~ x * z1 + z2, data = df)
```

Example workflow in R

```
m1 <- lm(y ~ x, data = df)
m2 <- lm(y ~ x + z1 + z2, data = df)
m3 <- lm(y ~ x * z1 + z2, data = df)
modelsummary(list(m1, m2, m3), vcov = "robust")
```

Example workflow in R

```
m1 <- lm(y ~ x, data = df)
m2 <- lm(y ~ x + z1 + z2, data = df)
m3 <- lm(y ~ x * z1 + z2, data = df)
modelsummary(list(m1, m2, m3), vcov = "robust")
modelplot(list(m1, m2, m3))
```

Example workflow in R

```
m1 <- lm(y ~ x, data = df)
m2 <- lm(y ~ x + z1 + z2, data = df)
m3 <- lm(y ~ x * z1 + z2, data = df)
modelsummary(list(m1, m2, m3), vcov = "robust")
modelplot(list(m1, m2, m3))
plot_predictions(m3, condition = c("x", "z1"))
```

Summary: Key takeaways

- Regression estimates conditional expectations

Summary: Key takeaways

- Regression estimates conditional expectations
- OLS is unbiased under standard assumptions

Summary: Key takeaways

- Regression estimates conditional expectations
- OLS is unbiased under standard assumptions
- Always use robust standard errors

Summary: Key takeaways

- Regression estimates conditional expectations
- OLS is unbiased under standard assumptions
- Always use robust standard errors
- Multiple regression: "holding constant" interpretation

Summary: Key takeaways

- Regression estimates conditional expectations
- OLS is unbiased under standard assumptions
- Always use robust standard errors
- Multiple regression: "holding constant" interpretation
- Control variables help only if chosen correctly

Summary: Key takeaways

- Regression estimates conditional expectations
- OLS is unbiased under standard assumptions
- Always use robust standard errors
- Multiple regression: "holding constant" interpretation
- Control variables help only if chosen correctly
- Interactions model conditional relationships

Summary: Key takeaways

- Regression estimates conditional expectations
- OLS is unbiased under standard assumptions
- Always use robust standard errors
- Multiple regression: "holding constant" interpretation
- Control variables help only if chosen correctly
- Interactions model conditional relationships
- Present results clearly: tables, coefficient plots, marginal effects

For next week

- Read Urdinez & Cruz (2020), chapter 8
- Read Gelman et al., chapters 13–14
- Complete Assignment 2

- Next session: Binary outcomes
  - → Linear probability model vs. logistic regression
  - → Interpreting logit results
  - → Predicted probabilities and marginal effects

Questions?