Welcome to Mastering Metrics!

Econ 140, Section 1

Jonathan Old

Roadmap

1. Getting Started & Housekeeping
2. Time for your questions
3. Econ 140: The Big Picture
4. Correlation and Causation

Getting Started & Housekeeping

Introduce yourself …

… and an interesting fact about you that you would like us to know!

Housekeeping

- What you can expect in this section and from me as GSI:

Housekeeping

- What you can expect in this section and from me as GSI:
  - Safe Space

Housekeeping

- What you can expect in this section and from me as GSI:
  - Safe Space
  - Diversity

Housekeeping

- What you can expect in this section and from me as GSI:
  - Safe Space
  - Diversity
  - Challenges in this course

Housekeeping

- What you can expect in this section and from me as GSI:
  - Safe Space
  - Diversity
  - Challenges in this course
  - Mental Health

Housekeeping

- What you can expect in this section and from me as GSI:
  - Safe Space
  - Diversity
  - Challenges in this course
  - Mental Health
- Core values: Transparency, Integrity, Respect

Housekeeping

- What you can expect in this section and from me as GSI:
  - Safe Space
  - Diversity
  - Challenges in this course
  - Mental Health
- Core values: Transparency, Integrity, Respect
- I am your learning companion. I see grades as feedback, not as an evaluation of your abilities or personality.

Your questions

Any questions?

… Remember – this is a safe space! Every question is useful!

Econ 140: The Big Picture

The Big Picture

- Why is applied econometrics so cool and so important?
- Why do we NOT use Excel? A cautionary tale
- Why econometrics and not data science?
    - Causality vs prediction
    - Economic models

Correlation and Causation

Dissecting Bad Causal Claims

Discuss in groups of 2: Why is this statement problematic?



Over the past 60 years, more spending on police hasn't meant less crime

The Washington Post

Intuitively, one might worry that reducing police spending would lead to a spike in crime. A review of spending on state and local police over the past 60 years, though, shows no correlation nationally between spending and crime rates.

In 1960, about $2 billion was spent by state and local governments on police. There were about 1,887 crimes per 100,000 Americans, including 161 violent crimes. By 1980, spending had increased to $14.6 billion — and crime rates had soared to 5,950 crimes per 100,000 Americans and 597 violent crimes. Over the next two decades, those rates thankfully fell, down to about 4,120 crimes per 100,000 people and 507 violent crimes. Spending spiked to more than $87 billion. Eighteen years later — by 2018, the most recent year for which full data are available — crime rates had fallen further to 2,580 crimes per 100,000, including 381 violent crimes.

Figure 1: Police spending and Crime (Source)

Dissecting Bad Causal Claims II

Discuss in groups of 2: Why is this statement problematic?

Three's a crowd: Having more than 2 kids linked to weaker brain function

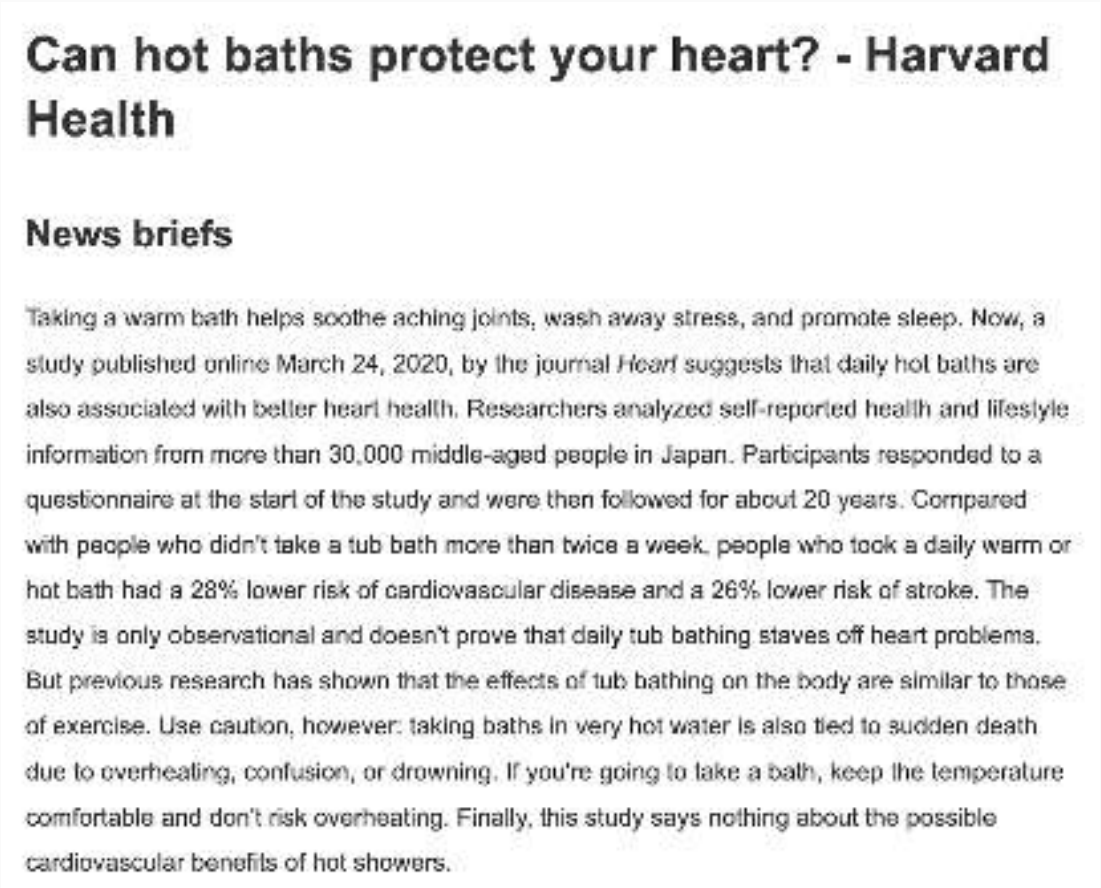


NEW YORK – Everything in moderation – even kids? Researchers from Columbia University and Université Paris-Dauphine report having more than two kids may have a negative impact on late-life cognition. The study shows that older parents with just two children appeared sharper cognitively than those with three.

Figure 2: Number of Children and Cognitive Function(Source)

Dissecting Bad Causal Claims III

Discuss in groups of 2: Why is this statement problematic?



Can hot baths protect your heart? - Harvard Health

News briefs

Taking a warm bath helps soothe aching joints, wash away stress, and promote sleep. Now, a study published online March 24, 2020, by the journal *Heart* suggests that daily hot baths are also associated with better heart health. Researchers analyzed self-reported health and lifestyle information from more than 30,000 middle-aged people in Japan. Participants responded to a questionnaire at the start of the study and were then followed for about 20 years. Compared with people who didn't take a tub bath more than twice a week, people who took a daily warm or hot bath had a 28% lower risk of cardiovascular disease and a 26% lower risk of stroke. The study is only observational and doesn't prove that daily tub bathing staves off heart problems. But previous research has shown that the effects of tub bathing on the body are similar to those of exercise. Use caution, however: taking baths in very hot water is also tied to sudden death due to overheating, confusion, or drowning. If you're going to take a bath, keep the temperature comfortable and don't risk overheating. Finally, this study says nothing about the possible cardiovascular benefits of hot showers.

Figure 3: Hot baths and health (Source)

Dissecting Bad Causal Claims IV

Discuss in groups of 2: Why is this statement problematic?



The New York Times

Another Benefit to Going to Museums? You May Live Longer

Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.

Leonardo da Vinci's "The Virgin and Child" at the Louvre in Paris. A British study found that people who engaged with the arts had a lower likelihood of dying during the study period than those who did not. Thibault Camus/Associated Press

Figure 4: Museums and longevity (Source)

Dissecting Bad Causal Claims IV

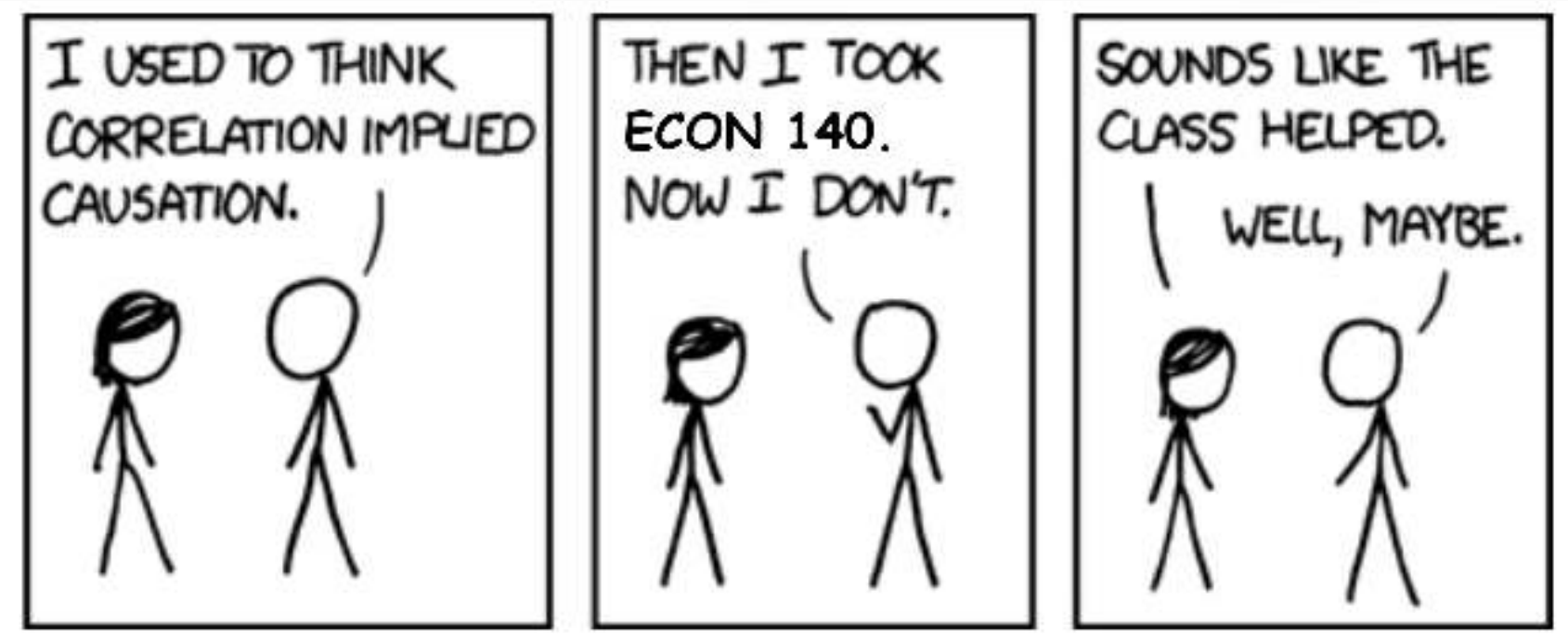


Figure 5: Correlation and causation

Quiz Time!

We have seen that
“Correlation Does Not Imply Causation”
What about:
“No Correlation Implies No Causation”?

Counterfactuals, Clones, and Cool Regression Stuff

Econ 140, Section 2

Jonathan Old

Roadmap

1. Potential outcomes

2. Selection bias

3. How to get rid of selection bias

Any questions?

… Remember – Every question is useful!

Potential outcomes

Potential outcomes

- "Potential outcomes" is a framework that can help us **think through causal claims**: An alternative to math, drawing errors, or thinking things through
- We like to **write things down** in a rigorous way: Transparent, easy to verify, easy to replicate
- Potential outcomes are hypothetical outcomes
- Example: Your exam score when you go to all sections vs. when you go to no sections
- Think of potential outcomes as "parallel universes"

Our main challenge: We NEVER observe an individual at more than one status at the same time!
$\rightarrow$ "The **fundamental problem of unobservability**"!

Potential outcomes: Notation

We write the potential outcomes as:

$$\left.\begin{aligned} Y_{i0} &= \text{Outcome of individual } i \text{ with "status" } 0 \\ Y_{i1} &= \text{Outcome of individual } i \text{ with "status" } 1 \end{aligned}\right\} \text{ Counterfactual Outcomes}$$

Alternative way of writing it: $Y_i(0)$ and $Y_i(1)$

"Status" can be anything

- Treatment assignment: 0 or 1
- Actual treatment: 0 or 1
- Drinking expensive whiskey or not
- Can also be: Multi-valued (number of children) or continuous (hours studied)

Potential outcomes: Notation

- We are often interested in the expected (think: average) potential outcome of a group of individuals with a given status.
- We write the group behind a conditional sign: $\mathrm{E}\left[\mathrm{Score}_{i0} \mid \mathrm{iPad}_i = 0\right]$ gives the potential outcome of a group of people that had no iPad, in the ”parallel universe” where they don’t have an iPad.
- Then, $\mathrm{E}\left[\mathrm{Score}_{i1} \mid \mathrm{iPad}_i = 0\right]$ gives the potental outcome of the same group (that currently have no iPad), in the ”parallel universe” where do have an iPad.

Treatment effects

We mostly use the potential outcomes framework to guide our thinking when we talk about estimating ”treatment effects”

- The individual treatment effect for agent i is: $Y_i(1) - Y_i(0)$
- The average treatment effect (ATE) is the expected value of the individual treatment effects: $ATE = E[Y_i(1) - Y_1(0)]$
- The Averae Treatment Effect on the Treated (ATT) is the same thing but conditioning on being treated: $ATT = E[Y_i(1) - Y_1(0)|T_i = 1]$
- Is the ATE always the same as the ATT?

Treatment effects

We mostly use the potential outcomes framework to guide our thinking when we talk about estimating ”treatment effects”

- The individual treatment effect for agent i is: $Y_i(1) - Y_i(0)$
- The average treatment effect (ATE) is the expected value of the individual treatment effects: $ATE = E[Y_i(1) - Y_1(0)]$
- The Averae Treatment Effect on the Treated (ATT) is the same thing but conditioning on being treated: $ATT = E[Y_i(1) - Y_1(0)|T_i = 1]$
- Is the ATE always the same as the ATT? **No!** It can be different from the ATE because the treated individuals need not be a random sample of the whole population.

Estimating the effect of iPads on grades

Let us start with a difference-in-means comparison:

$$\Delta = E[\text{Grade}_i|\text{iPad}_i = 1] - E[\text{Grade}_i|\text{iPad}_i = 0]$$

Add and subtract $E[\text{Grade}_i(0)|\text{iPad}_i = 1]$:

$$= E[\text{Grade}_i(1)|\text{iPad}_i = 1] - E[\text{Grade}_i(0)|\text{iPad}_i = 1] \quad +$$
$$E[\text{Grade}_i(0)|\text{iPad}_i = 1] - E[\text{Grade}_i(0)|\text{iPad}_i = 0]$$

Use properties of expectations:

$$= E[\text{Grade}_i(1) - \text{Grade}_i(0)|\text{iPad}_i = 1] \quad +$$
$$E[\text{Grade}_i(0)|\text{iPad}_i = 1] - E[\text{Grade}_i(0)|\text{iPad}_i = 0]$$
$$= \text{ATT} + \text{Selection bias}$$

Selection bias: Students with and without iPad have different potential grades: **even if they both *had* iPads, they would be different**.

Dissecting Bad Causal Claims

Discuss in groups of 2: Why is this statement problematic?



Over the past 60 years, more spending on police hasn't meant less crime

The Washington Post

Intuitively, one might worry that reducing police spending would lead to a spike in crime. A review of spending on state and local police over the past 60 years, though, shows no correlation nationally between spending and crime rates.

In 1960, about $2 billion was spent by state and local governments on police. There were about 1,887 crimes per 100,000 Americans, including 161 violent crimes. By 1980, spending had increased to $14.6 billion — and crime rates had soared to 5,950 crimes per 100,000 Americans and 597 violent crimes. Over the next two decades, those rates thankfully fell, down to about 4,120 crimes per 100,000 people and 507 violent crimes. Spending spiked to more than $87 billion. Eighteen years later — by 2018, the most recent year for which full data are available — crime rates had fallen further to 2,580 crimes per 100,000, including 381 violent crimes.

Figure 1: Police spending and Crime (Source)

Dissecting Bad Causal Claims II

Discuss in groups of 2: Why is this statement problematic?

Three's a crowd: Having more than 2 kids linked to weaker brain function



NEW YORK – Everything in moderation – even kids? Researchers from Columbia University and Université Paris-Dauphine report having more than two kids may have a negative impact on late-life cognition. The study shows that older parents with just two children appeared sharper cognitively than those with three.

Figure 2: Number of Children and Cognitive Function(Source)

Dissecting Bad Causal Claims III

Discuss in groups of 2: Why is this statement problematic?



The New York Times

Another Benefit to Going to Museums? You May Live Longer

Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.

Leonardo da Vinci's "The Virgin and Child" at the Louvre in Paris. A British study found that people who engaged with the arts had a lower likelihood of dying during the study period than those who did not. Thibault Camus/Associated Press

Figure 3: Museums and longevity (Source)

Selection bias

How to think about Selection Bias

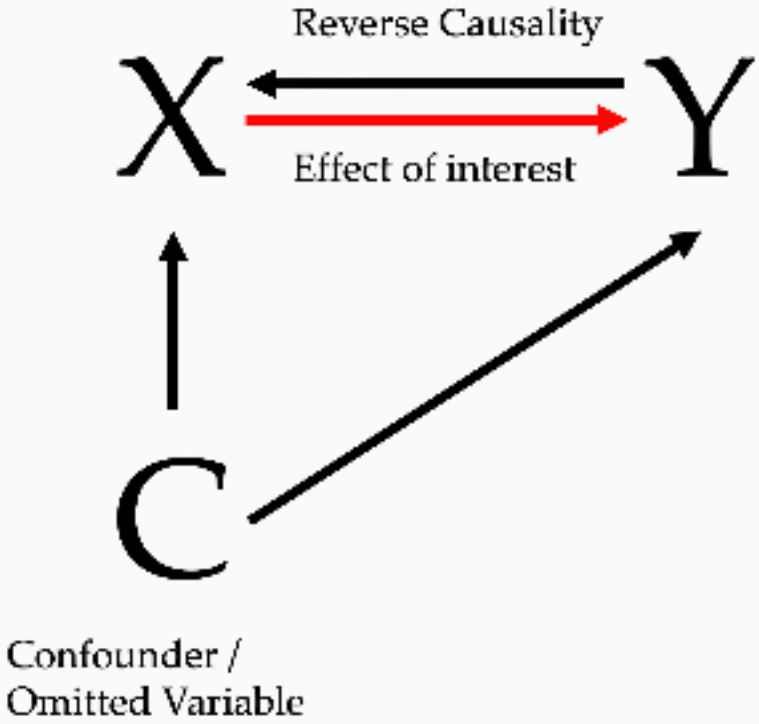


Figure 4: Selection bias

How to get rid of selection bias

RCTs solve selection bias

We had:

$$\Delta = E[\text{Grade}_i(1) - \text{Grade}_i(0)|\text{iPad}_i = 1] \quad + \\ E[\text{Grade}_i(0)|\text{iPad}_i = 1] - E[\text{Grade}_i(0)|\text{iPad}_i = 0]$$

- The second line was selection bias: The potential grade of individuals with and without iPad is different
- If the treatment (iPad) is independent of the potential outcomes, then:

$$\text{iPad}_i \perp (\text{Grade}_i(1), \text{Grade}_i(0))$$

$$\Rightarrow E[\text{Grade}_i(0)|\text{iPad}_i = 1] = E[\text{Grade}_i(0)|\text{iPad}_i = 0]$$

  and selection bias will be zero.
- The the difference is equal to the ATT and **also the ATE**:

$$\Delta = E[\text{Grade}_i(1) - \text{Grade}_i(0)|\text{iPad}_i = 1] \\ = E[\text{Grade}_i(1) - \text{Grade}_i(0)|\text{iPad}_i = 0]$$

Regression in Action: Bias, squares, and measurement error

Econ 140, Section 3

Jonathan Old

Roadmap

1. Recap
2. Selection bias in action: Omitted Variable Bias
3. Quadratic Terms
4. Measurement Error

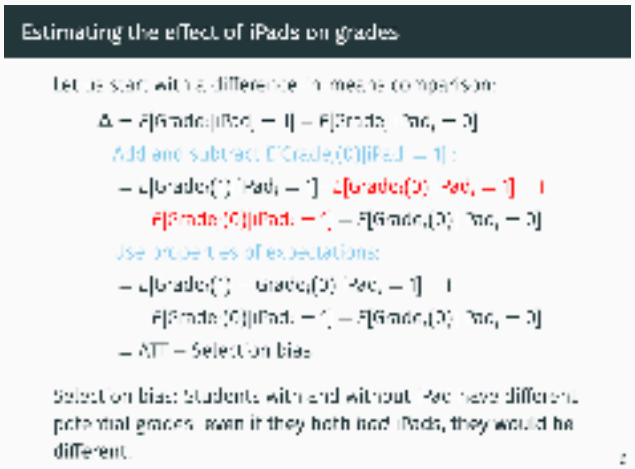
Any questions?

… Remember – Every question is useful!

Recap

Recap: Selection bias

We saw that whenever we do a difference-in-means comparison (or a regression), we get:



Causal Effect

+

Selection Bias

- We cannot observe them - so we can never be sure!
- Econometrics is all about uncertainty. You can always state that there are different possibilities and you cannot know for sure

Recap: How to think about Selection Bias

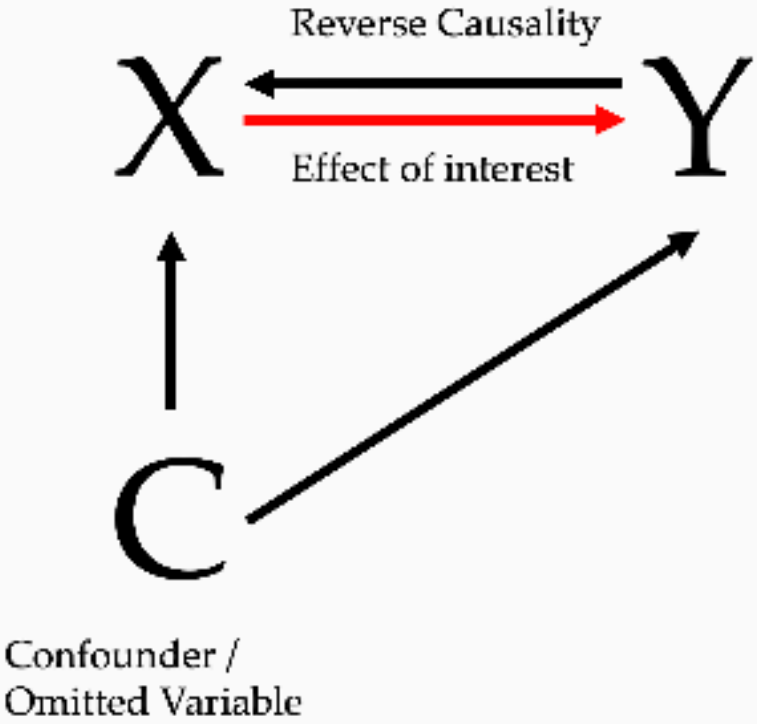


Figure 1: Selection bias

Selection bias in action: Omitted Variable Bias

Omitted Variable Bias (OVB)

Let's go through an example:

You are doing a summer internship with San-Francisco based tech company "Chirp", when your boss asks you to help evaluating the effect of "performance-based" or "incentive" pay on employees' job satisfaction. When you analyze the data, you realize that it's a mess. Until now, every manager could decide on the performance pay system in their team. Can you run a regression of job satisfaction on whether or not an individual works in a team that has a performance bonus in place?

You believe that:

$$\text{Satisfaction}_i = \beta_0 + \beta_1 \text{Performance pay}_i + \beta_2 \text{Manager quality}_i + u_i$$

$$\text{Manager quality}_i = \gamma_0 + \gamma_1 \text{Performance pay}_i + v_i$$

OVB Question

1. Run a ”short” regression of satisfaction (S_i) on pay (P_i):

$$S_i = \alpha_0 + \alpha_1 P_i + e_i$$

OVB Question

1. Run a "short" regression of satisfaction (S_i) on pay (P_i):

$$S_i = \alpha_0 + \alpha_1 P_i + e_i$$

$$\alpha_1 = \frac{\text{Cov}(S_i, P_i)}{\text{Var}(P_i)}$$

2. Solve this expression with what we know

OVB Question

1. Run a "short" regression of satisfaction (S_i) on pay (P_i):

$$S_i = \alpha_0 + \alpha_1 P_i + e_i$$

$$\alpha_1 = \frac{\text{Cov}(S_i, P_i)}{\text{Var}(P_i)}$$

2. Solve this expression with what we know

$$\alpha_1 = \frac{\text{Cov}(S_i, P_i)}{\text{Var}(P_i)}$$

OVB Question

1. Run a "short" regression of satisfaction (S_i) on pay (P_i):

$$\mathrm{S}_i = \alpha_0 + \alpha_1 \mathrm{P}_i + e_i$$

$$\alpha_1 = \frac{\mathrm{Cov}(S_i, P_i)}{\mathrm{Var}(P_i)}$$

2. Solve this expression with what we know

$$\alpha_1 = \frac{\mathrm{Cov}(S_i, P_i)}{\mathrm{Var}(P_i)} = \frac{\mathrm{Cov}(\beta_0 + \beta_1 \mathrm{P}_i + \beta_2 \mathrm{M}_i + u_i, P_i)}{\mathrm{Var}(P_i)}$$

OVB Question

1. Run a "short" regression of satisfaction (S_i) on pay (P_i):

$$S_i = \alpha_0 + \alpha_1 P_i + e_i$$

$$\alpha_1 = \frac{\text{Cov}(S_i, P_i)}{\text{Var}(P_i)}$$

2. Solve this expression with what we know

$$\alpha_1 = \frac{\text{Cov}(S_i, P_i)}{\text{Var}(P_i)} = \frac{\text{Cov}(\beta_0 + \beta_1 P_i + \beta_2 M_i + u_i, P_i)}{\text{Var}(P_i)}$$

$$= \beta_1 \underbrace{\frac{\text{Cov}(P_i, P_i)}{\text{Var}(P_i)}}_{=1} + \beta_2 \underbrace{\frac{\text{Cov}(M_i, P_i)}{\text{Var}(P_i)}}_{=\gamma_1} + \underbrace{\frac{\text{Cov}(u_i, P_i)}{\text{Var}(P_i)}}_{=0, \text{ by assumption}}$$

OVB Question

1. Run a "short" regression of satisfaction (S_i) on pay (P_i):

$$S_i = \alpha_0 + \alpha_1 P_i + e_i$$

$$\alpha_1 = \frac{\text{Cov}(S_i, P_i)}{\text{Var}(P_i)}$$

2. Solve this expression with what we know

$$\alpha_1 = \frac{\text{Cov}(S_i, P_i)}{\text{Var}(P_i)} = \frac{\text{Cov}(\beta_0 + \beta_1 P_i + \beta_2 M_i + u_i, P_i)}{\text{Var}(P_i)}$$

$$= \beta_1 \underbrace{\frac{\text{Cov}(P_i, P_i)}{\text{Var}(P_i)}}_{=1} + \beta_2 \underbrace{\frac{\text{Cov}(M_i, P_i)}{\text{Var}(P_i)}}_{=\gamma_1} + \underbrace{\frac{\text{Cov}(u_i, P_i)}{\text{Var}(P_i)}}_{=0, \text{ by assumption}}$$

$$= \underbrace{\beta_1}_{\text{Actual effect}} + \underbrace{\beta_2 \gamma_1}_{\text{Omitted variable bias}}$$

The most important slide of this course (until the midterm)

We can summarize everything of OVB in three equations. Let Y_i be the outcome variable, X_i our regressor of interest, and Z_i the "omitted" variable.

$$\begin{aligned} \text{[Long regression]} \quad & Y_i = c_1 + \beta_L X_i + \delta Z_i + e_i \\ \text{[Short regression]} \quad & Y_i = c_2 + \beta_S X_i + u_i \\ \text{[Auxiliary regression]} \quad & Z_i = c_3 + \gamma X_i + v_i \end{aligned}$$

Then, the **Omitted variable bias formula** states that:

$$\underbrace{\beta_S}_{\text{Short}} = \underbrace{\beta_L}_{\text{Long}} + \underbrace{\delta}_{\text{Omitted}} \cdot \underbrace{\gamma}_{\text{Included}}$$

We call $\delta\gamma$ the **omitted variable bias**. We can appraise the direction of the bias by multiplying our guesses for the signs of δ and γ.

Wait a minute!

We saw that the OLS coefficient in the short regression is:

$$\alpha_1 = \underbrace{\beta_1}_{\text{Actual effect}} + \underbrace{\beta_2\gamma_1}_{\text{Omitted variable bias}}$$

This looks awfully much like:

Estimating the effect of iPads on grades

Let us start with a difference-in-means comparison:

$$\Delta = E[\text{Grade}_i|\text{iPad}_i = 1] - E[\text{Grade}_i|\text{iPad}_i = 0]$$

Add and subtract $E[\text{Grade}_i(0)|\text{iPad}_i = 1]$:

$$= E[\text{Grade}_i(1)|\text{iPad}_i = 1] - E[\text{Grade}_i(0)|\text{iPad}_i = 1] + E[\text{Grade}_i(0)|\text{iPad}_i = 1] - E[\text{Grade}_i(0)|\text{iPad}_i = 0]$$

Use properties of expectations:

$$= E[\text{Grade}_i(1) - \text{Grade}_i(0)|\text{iPad}_i = 1] + E[\text{Grade}_i(0)|\text{iPad}_i = 1] - E[\text{Grade}_i(0)|\text{iPad}_i = 0]$$

$$= \text{ATT} + \text{Selection bias}$$

Selection bias: Students with and without iPad have different potential grades: **even if they both *had* iPads, they would be different.**

5

Control variables

Control variables are additional variables (or covariates) **included in a regression**. We do this for **various reasons** (in decreasing order of importance):

- To remove selection bias / omitted variable bias
- To increase precision of our estimates
- To know about the (conditional/partial) correlation of other variables
- To better predict the outcome

Let's see graphically how control variables work.

Any questions?

… Remember – Every question is useful!

Attendance and Evaluation

Please fill out a very short evaluation form here:
https://forms.gle/7neuAJbkAv17mSEp6



Quadratic Terms

Quadratic Terms

See in *RStudio*

Measurement Error

Measurement Error

See on Blackboard and in *RStudio*.

If time permits, see on chatGPT.

Getting fit for the Midterm!

Econ 140, Section 4

Jonathan Old

Roadmap

1. Recap
2. Interaction terms (Q6)
3. Logs (Q4)
4. Topics we've glossed over so far

Any questions?

… Some comments on the evaluations asked for more space to answer left-over questions from the lecture: Now is the time!

Recap

Recap: OVB (Very important!)

We can summarize everything of OVB in three equations. Let Y_i be the outcome variable, X_i our regressor of interest, and Z_i the "omitted" variable.

$$\begin{aligned} \text{[Long regression]} \quad & Y_i = c_1 + \beta_L X_i + \delta Z_i + e_i \\ \text{[Short regression]} \quad & Y_i = c_2 + \beta_S X_i + u_i \\ \text{[Auxiliary regression]} \quad & Z_i = c_3 + \gamma X_i + v_i \end{aligned}$$

Then, the **Omitted variable bias formula** states that:

$$\underbrace{\beta_S}_{\text{Short}} = \underbrace{\beta_L}_{\text{Long}} + \underbrace{\delta}_{\text{Omitted}} \cdot \underbrace{\gamma}_{\text{Included}}$$

We call $\delta\gamma$ the **omitted variable bias**. We can appraise the direction of the bias by multiplying our guesses for the signs of δ and γ. **If either $\delta = 0$ or $\gamma = 0$, then OVB is zero!**

Recap: Understanding bias in OLS regressions

- We can use the OLS formula to understand how bias works in OLS regression

$$\hat{\beta}_1 = \frac{\mathsf{Cov}(X_i, Y_i)}{\mathsf{Var}(X_i)}$$

- For **OVB**: We *know* the true Y_i and plug it in
- For **measurement error**: We **know** what X_i and plug it in
- Simplify using the following rules:

  $\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y) + 2\,\mathsf{Cov}(X, Y)$

  $\mathsf{Var}(X - Y) = \mathsf{Var}(X) + \mathsf{Var}(Y) - 2\,\mathsf{Cov}(X, Y)$

  $\mathsf{Cov}(aX + bY, Z) = a\,\mathsf{Cov}(X, Z) + b\,\mathsf{Cov}(Y, Z)$

  $\mathsf{Cov}(X, X) = \mathsf{Var}(X)$

  $\mathsf{Var}(aX) = a^2\,\mathsf{Var}(X)$

  $\mathsf{Cov}(X, Y) = 0$, if X and Y are independent.

  $\mathsf{Var}(X) \geq 0$.

Recap: Making OLS more interesting

- We saw that we can extend the simple OLS framework

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

  to something richer:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i$$

- We will get to know many more versions of this today
- All questions of the type *"how is Y_i expected to change if we change X_i"* can be solved with **partial derivatives** – in this case:

$$\frac{\partial Y_i}{\partial X_i} =$$

Recap: Making OLS more interesting

- We saw that we can extend the simple OLS framework

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

  to something richer:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i$$

- We will get to know many more versions of this today
- All questions of the type *"how is Y_i expected to change if we change X_i"* can be solved with **partial derivatives** – in this case:

$$\frac{\partial Y_i}{\partial X_i} = \beta_1 + 2 \cdot \beta_2 \cdot X_i$$

Interaction terms (Q6)

Interaction terms: Making OLS more interesting (Q6)

Let us consider the model

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

where Y_i is a country's GDP per capita, X_{1i} the value of its natural resources, and X_{2i} a measure of how democratic it is.

1. How do we interpret β_1?

Interaction terms: Making OLS more interesting (Q6)

Let us consider the model

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

where Y_i is a country's GDP per capita, X_{1i} the value of its natural resources, and X_{2i} a measure of how democratic it is.

1. How do we interpret β_1?
   Keeping democracy fixed, increasing the value of a country's natural resources by one unit is associated with β_1 higher GDP per capita.
2. How do we interpret β_2?

Interaction terms: Making OLS more interesting (Q6)

Let us consider the model

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

where Y_i is a country's GDP per capita, X_{1i} the value of its natural resources, and X_{2i} a measure of how democratic it is.

1. How do we interpret β_1?
   Keeping democracy fixed, increasing the value of a country's natural resources by one unit is associated with β_1 higher GDP per capita.
2. How do we interpret β_2?
   Keeping natural resources fixed, increasing a country's democracy score by one unit is associated with β_2 higher GDP per capita.

Interaction Terms (ii) (Q6)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the ”effect” of X_{1i} on Y_i?

Interaction Terms (ii) (Q6)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the ”effect” of X_{1i} on Y_i? $\beta_1 + \beta_3 X_{2i}$

Interaction Terms (ii) (Q6)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the "effect" of X_{1i} on Y_i? $\beta_1 + \beta_3 X_{2i}$
2. How do we interpret β_1?

Interaction Terms (ii) (Q6)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the "effect" of X_{1i} on Y_i? $\beta_1 + \beta_3 X_{2i}$
2. How do we interpret β_1? **The effect of an additional unit of X_{1i} , if X_{2i} is equal to 0.**

Interaction Terms (ii) (Q6)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the "effect" of X_{1i} on Y_i? $\beta_1 + \beta_3 X_{2i}$
2. How do we interpret β_1? **The effect of an additional unit of X_{1i} , if X_{2i} is equal to 0.**
3. How do we interpret β_2?

Interaction Terms (ii) (Q6)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the "effect" of X_{1i} on Y_i? $\beta_1 + \beta_3 X_{2i}$
2. How do we interpret β_1? **The effect of an additional unit of X_{1i} , if X_{2i} is equal to 0.**
3. How do we interpret β_2? **The effect of an additional unit of X_{2i} , if X_{1i} is equal to 0.**

Interaction Terms (ii) (Q6)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the "effect" of X_{1i} on Y_i? $\beta_1 + \beta_3 X_{2i}$
2. How do we interpret β_1? **The effect of an additional unit of X_{1i} , if X_{2i} is equal to 0.**
3. How do we interpret β_2? **The effect of an additional unit of X_{2i} , if X_{1i} is equal to 0.**
4. How do we interpret $\beta_1 + \beta_3$?

Interaction Terms (ii) (Q6)

Now, let us extend the model to:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i$$

1. What is the "effect" of X_{1i} on Y_i? $\beta_1 + \beta_3 X_{2i}$
2. How do we interpret β_1? **The effect of an additional unit of X_{1i} , if X_{2i} is equal to 0.**
3. How do we interpret β_2? **The effect of an additional unit of X_{2i} , if X_{1i} is equal to 0.**
4. How do we interpret $\beta_1 + \beta_3$? **The effect of an additional unit of X_{1i} , if X_{2i} is equal to 1.**

Rule of thumb: Always use partial derivatives to make sure that you are right!

Practice exam question: 1a)

The Ministry of Truth is interested in a rumour that **air pollution could impact mental health**. One of the most harmful pollutants is fine particulate matter PM2.5, which comes from operations that involve the burning of fuels such as wood, oil, coal, etc. A research team is sent to investigate the rumour. The team **randomly** selects and surveys 19,920 people across 71 districts of the country. The key variable, Exposure E_i, is a **dummy variable** equal to 1 if the individual i is exposed to a large amount of PM2.5 in the last 24 hours, and 0 otherwise. The team also conducts a standardised questionnaire to record **depressive symptoms** in the last month, called the Kessler Psychological Distress scale (K6). The questionnaire results in a score, Depression_i, that ranges from 0 to 24; and the higher the score, the more severe the depressive symptoms for individual i. The variable has a sample average of 2.96. **Running regressions with Depression D_i as the dependent variable, you obtain the following results**:

Practice exam question: 1a)

Dependent variable: Depression $_i$

Regressor	(1)	(2)	(3)
Exposure $_i$	0.834	0.614	0.554
	(0.032)	(0.045)	(0.042)
Exposure $_i \times$ Female $_i$		0.065	
		(0.024)	
Female $_i$		−0.739	−0.825
		(0.036)	(0.066)
Age_i			0.452
			(0.132)
Age_i^2			0.524
			(0.121)

Notes: All estimations contain a constant term. Robust standard errors are in the parentheses. Age_i is the age (years old) of individual i, and Age_i^2 is the square of Age_i.

Pratice Exam question: 1a)

a) Interpreting the coefficient in Column (1), a journalist, Katherine, claims: ”Since participants are randomly selected, we can infer that exposure to a large amount of PM2.5 does cause depression.”

i. Explain carefully why Katherine is wrong, specifying the direction of bias(es) if there is any. Which assumption(s) would she need to impose for the causality claim to hold?

ii. What is the correct interpretation from Column (1) that Katherine should have made?

(Detailed) Suggested Answer: 1a)

i. **Random selection is not the same thing as random assignment to treatment!** Survey respondents may be systematically different from each other in ways that are correlated with depression and pollution exposure. Therefore, the results from a regression can not be interpreted causally (and are biased). A priori, it is unclear in which direction the bias would go, but we could imagine that (**Only one explanation needed for exam**):
On rainy days, pollution is lower (-) and people may be reporting more depression symptoms (+), leading to downward bias. More wealthy people choose to live in less polluted areas (-) and they may have less depression (e.g., better access to mental health resources) (-), leading to upward bias.
For the regression causality claim to hold, we need to assume that: People exposed to pollution and those not exposed to pollution would have, on average, the same depression level, had they been exposed to the same level of pollution. In other words: **Both groups would have to have the same potential depression outcomes**.

ii. On average, people that were exposed to pollution had a 0.8 points higher score on the depression scale. The difference between the two groups is significant at the 5% level.

Pratice Exam question: 1b)

b) Interpret column (2) of the regression table

i. A colleague notes the the coefficient on Female_i is significant, and states: ”The effect of being female on depression is significantly different from zero”. Do you agree with the statement? Why or why not?

ii. How is pollution exposure related to depression, for men? And how for women?

(Detailed) Suggested Answer: 1b)

i. It is difficult to make such interpretations when interaction terms are involved. Taking partial derivatives, the "effect" of being female is:

$$\frac{\partial \text{Depression}_i}{\partial \text{Female}_i} = -0.739 + 0.065 \cdot \text{Exposure}_i$$

We can do inference (and test significance) at $\text{Exposure}_i = 0$ (just looking at the coefficient for female, -0.739 is significant). We can also do it for any other level of exposure, but for that we also need to take the other coefficient into account and cannot just use the table.

Comment 1: Less relevant for the exam, but important to be aware of!
Comment 2: We can always do inference on the interaction term, which is significant here!

ii. Taking partial derivatives, the "effect" of pollution exposure is:

$$\frac{\partial \text{Depression}_i}{\partial \text{Exposure}_i} = 0.614 + 0.065 \cdot \text{Female}_i$$

Hence, the "effect" for Males is 0.614 and the "effect" for Females is larger ($0.614 + 0.065 = 0.779$). The effect of pollution is also **signficantly** larger than for men, because the coefficient on the interaction term is significantly different from zero.

Logs (Q4)

Notes on logarithms (Q4)

- We can take logs of whole equations to get linear models (problem set)
- We can also take logs of specific variables, especially when they have long tails (wealth in the US, GDP per capita, etc.)
- We can get to the right interpretation of log-specifications with just math

Notes on logarithms (Q4)

- We can take logs of whole equations to get linear models (problem set)
- We can also take logs of specific variables, especially when they have long tails (wealth in the US, GDP per capita, etc.)
- We can get to the right interpretation of log-specifications with just math
- But I will make your life easier with a cheat sheet.

Logs: Cheatsheet (Wooldridge version) (Q4)

Summary of Functional Forms Involving Logarithms

Model	LHS	RHS	Interpretation of β_1
Level-level	y	x	$\Delta y = \beta_1 \Delta x$
Level-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\,\%\Delta x$
Log-level	$\log(y)$	x	$\%\Delta y = (100\beta_1)\,\Delta x$
Log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

Table taken from Wooldridge (2011)

Logs: Cheatsheet II (Jonathan's version) (Q4)

Model	LHS	RHS	A change in x by ...	is associated with a change in y by ...
Level-Level	y	x	1 unit	β_1 units
Level-Log	y	$\log(x)$	1%	$\beta_1/100$ units
Log-Level	$\log(y)$	x	1 unit	$100\beta_1$%
Log-Log	$\log(y)$	$\log(x)$	1%	β_1%

If you want to get a bonus star from me, write "approximately" in log-interpretations.

Topics we've glossed over so far

Any questions?

… Some comments on the evaluations asked for more space to answer left-over questions from the lecture: Now is the time!

Bad controls

- Not all controls are good controls
- Some controls are called ”bad controls”. These are:
  1. Variables that are themselves outcomes of a treatment: What happens if you control for the change in English test scores in the regression below?

	Treatment	Control
Change in Math Scores	2	1
Change in English Scores	2	1

  2. Variables that moderate the treatment effect, e.g. controlling for occupation choice in gender wage gap regression . . .
- Rule of Thumb: Good controls are either pre-determined or immutable characteristics.
- Another way to think about it: Controls help us make ”apples to apples” comparisons. Which apples matter?

What if the outcome variable is binary (a dummy variable)?

Let's run the regression

$$\text{Defaulted}_i = \alpha + \beta \widetilde{\text{Credit Score}}_i + e_i$$

where Defaulted_i is equal to 1 if individual i has ever defaulted on a loan (mortgage, credit card, auto loan, etc.), and $\widetilde{\text{Credit Score}}_i$ is i's credit score, **minus the average credit score in the sample** (Note: US credit scores range from 300 to 850 points).

1. You run a regression and get $\hat{\alpha}$=0.1. How do you interpret this? Does this number make sense here?
2. Your estimate for β is $\hat{\beta} = 0.001$. Interpret.

What if the outcome variable is binary (a dummy variable)?

Let's run the regression

$$\text{Defaulted}_i = \alpha + \beta \tilde{\text{Credit Score}}_i + e_i$$

where Defaulted_i is equal to 1 if individual i has ever defaulted on a loan (mortgage, credit card, auto loan, etc.), and $\tilde{\text{Credit Score}}_i$ is i's credit score, **minus the average credit score in the sample** (Note: US credit scores range from 300 to 850 points).

1. You run a regression and get $\hat{\alpha}$=0.1. How do you interpret this? Does this number make sense here?
2. Your estimate for β is $\hat{\beta} = 0.001$. Interpret.

With a dummy dependent variable, changing X_i by one unit increases the probability of $Y_i = 1$ by $\hat{\beta} \cdot 100$ percentage points.

Inference and the variance of $\hat{\beta}_{OLS}$ (Q3)

The variance of the OLS estimator is $\mathbf{Var}(\hat{\beta}_1^{OLS}) = \frac{\sigma_\epsilon^2}{N \cdot \mathbf{Var}(X_i)}$.
We expect to get more precise estimates if

- The variance of X_i increases
- The variance of the error term ϵ_i decreases
- The sample size N increases

Hypothesis testing

$$\left| \frac{\hat{\beta}}{\text{SE}(\hat{\beta})} \right| \geq 1.96$$

$$\Leftrightarrow \left| \text{t-stat} \right| \geq 1.96$$

$$\Leftrightarrow \text{p-value} \leq 0.05$$

If you are testing the null hypothesis H_0: $\beta = 0$, then all of these are equivalent, and you can use any of these.

The power of RCTs – Teacher Performance Pay in India

Econ 140, Section 6

Jonathan Old

Roadmap

1. Recap: Why RCTs? [SA7-Q1]
2. Teacher Performance Pay [SA7-Q2]
3. Conclusion
4. Other tables if there is extra time

Any comments about the midterm?

Recap: Why RCTs? [SA7-Q1]

RCTs solve selection bias [SA7-Q1a]

We had:

$$\Delta = E[\text{Score}_i(1) - \text{Score}_i(0)|\text{PPay}_i = 1] \quad + \\ E[\text{Score}_i(0)|\text{PPay}_i = 1] - E[\text{Score}_i(0)|\text{PPay}_i = 0]$$

- The second line was selection bias: The potential Score of individuals with and without Performance Pay is different
- If the treatment (PPay) is independent of the potential outcomes, then:

$$\text{PPay}_i \perp (\text{Score}_i(1), \text{Score}_i(0))$$

$$\Rightarrow E[\text{Score}_i(0)|\text{PPay}_i = 1] = E[\text{Score}_i(0)|\text{PPay}_i = 0]$$

and selection bias will be zero.

- The the difference is equal to the ATT and **also the ATE**:

$$\Delta = E[\text{Score}_i(1) - \text{Score}_i(0)|\text{PPay}_i = 1] \\ = E[\text{Score}_i(1) - \text{Score}_i(0)|\text{PPay}_i = 0]$$

RCTs and omitted variable bias [SA7-Q1b]

Two ways to think about this

1. Potential outcomes framework: We can just compare the differences in means, and we do not need to control for anything $\Rightarrow$ No OVB!
2. OVB formula:

$$\beta_S = \beta_L + \delta\gamma$$

.

- The bias is $\delta\gamma$, where γ measured the association between X (treatment) and the omitted variable
- But in an RCT, this association is zero $\Rightarrow$ No OVB!

Do RCTs solve everything? [SA7-Q1c]

Discuss in groups of 4: Why don't RCTs solve everything?

Your concerns may be:

- ethical
- practical
- remaining econometric challenges
- others

Do RCTs solve everything? [SA7-Q1c]

Discuss in groups of 4: Why don't RCTs solve everything?

Your concerns may be:

- ethical
- practical
- remaining econometric challenges
- others

Example from my life: Job interview

RCTs: Remaining challenges [SA7-Q1c]

- Non-compliance
- Differential attrition
- Spillover effects / general equilibrium effects
- Practical concerns
- Ethical concerns
- External validity (Generalizability)
- Placebo effects / experimenter demand effects / Hawthorne effect

RCTs have revolutionized economics

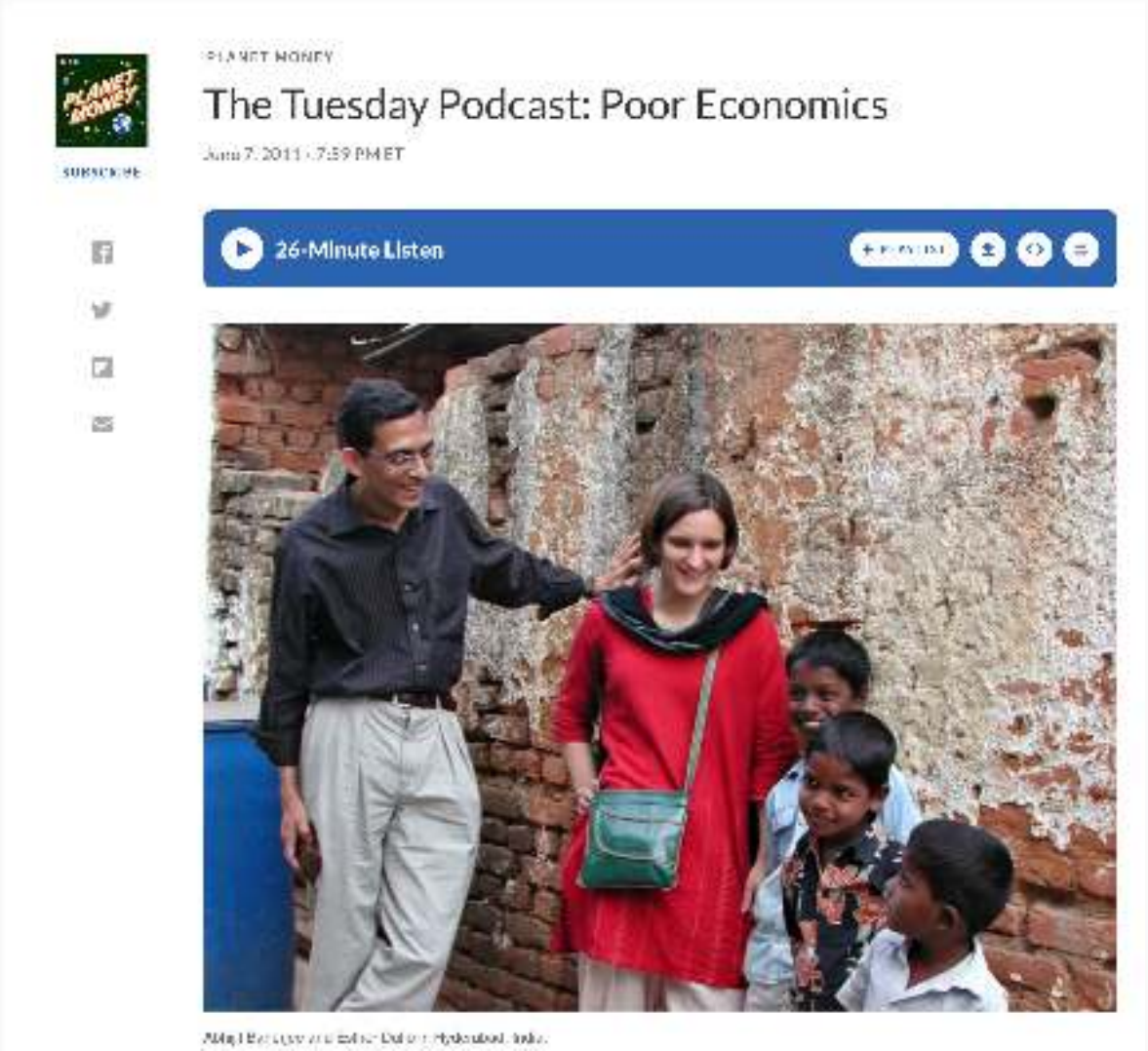
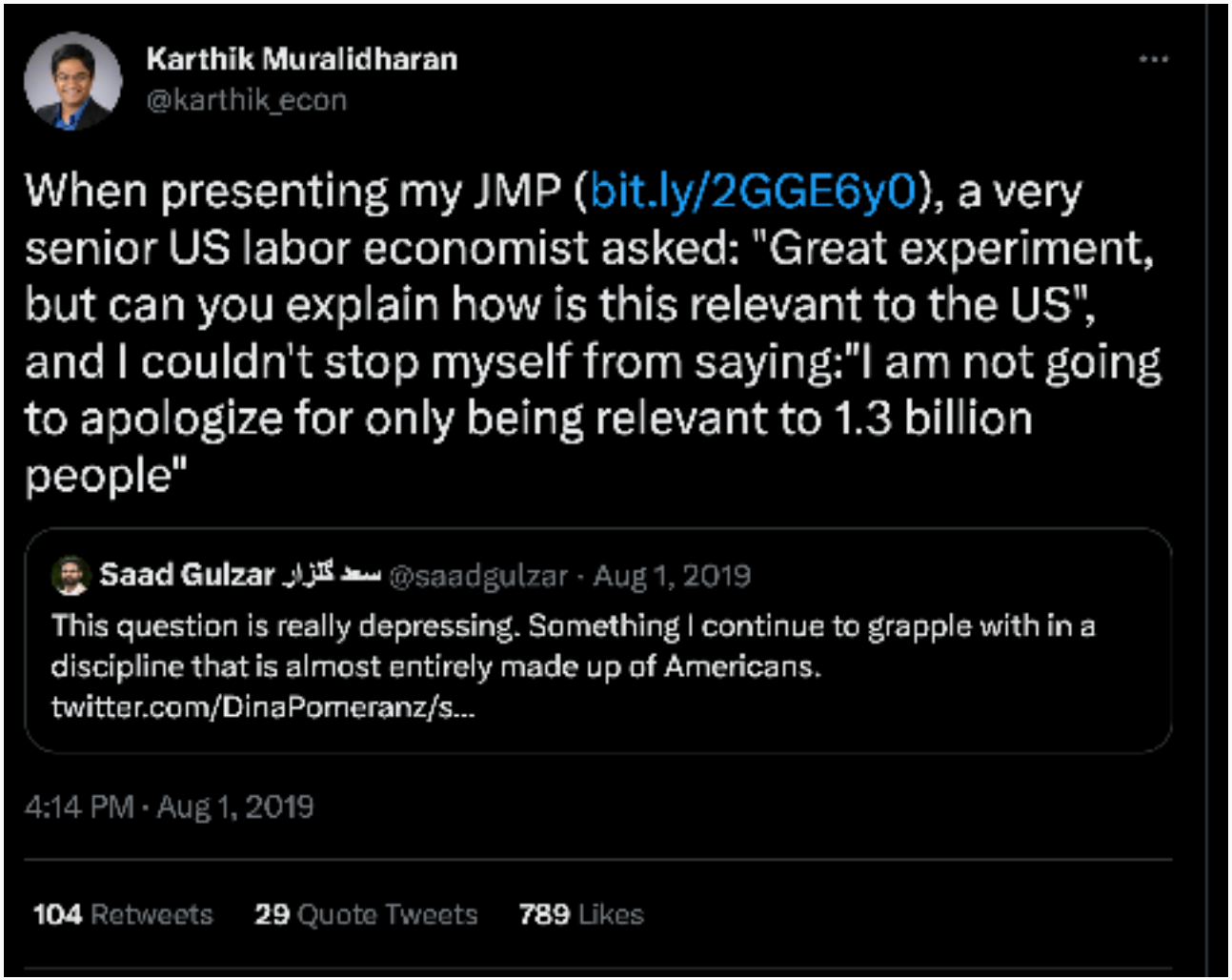


Figure 1: Abhijeet Banerjee and Ester Duflo. (Source)

Teacher Performance Pay [SA7-Q2]

Why does this matter?

Teacher absenteeism is a huge problem in many countries across the world!



Intervention [SA7-Q2b]

$$\text{Bonus} = \begin{cases} \text{Rs. } 500 \times (\%\text{ gain in average test scores} - 5\%) & \text{if gain} > 5\% \\ 0 & \text{otherwise.} \end{cases}$$

- How much bonus does a teacher get if their students' average test score changes by -10%, 0%, 5%, 10%, 50%?
- Is it ethical?
- What do you expect the effect to be? Which direction? Large or small? (teachers earn 180 USD without incentive)

Experimental Design [SA7-Q2b]

TABLE 1
INCENTIVES

	Incentives (Conditional on Improvement in Student Learning)		
Inputs	None	Group Bonus	Individual Bonus
None	Control (100 schools)	100 schools	100 schools
Extra contract teacher	100 schools		
Extra block grant	100 schools		

- Why randomize at the level of schools (not teachers?)

Experimental Design [SA7-Q2b]

TABLE 1
INCENTIVES

	INCENTIVES (Conditional on Improvement in Student Learning)		
INPUTS	None	Group Bonus	Individual Bonus
None	Control (100 schools)	100 schools	100 schools
Extra contract teacher	100 schools		
Extra block grant	100 schools		

- Why randomize at the level of schools (not teachers?)
  Much easier to implement, prevent undesirable effects within schools (spillovers, jealousy, etc.)

Experimental Design [SA7-Q2b]

TABLE 1
INCENTIVES

	Incentives (Conditional on Improvement in Student Learning)		
Inputs	None	Group Bonus	Individual Bonus
None	Control (100 schools)	100 schools	100 schools
Extra contract teacher	100 schools		
Extra block grant	100 schools		

- Why randomize at the level of schools (not teachers?)
  Much easier to implement, prevent undesirable effects within schools (spillovers, jealousy, etc.)
- What is the difference between the individual bonus and the group bonus?

Experimental Design [SA7-Q2b]

TABLE 1
INCENTIVES

	INCENTIVES (Conditional on Improvement in Student Learning)		
INPUTS	None	Group Bonus	Individual Bonus
None	Control (100 schools)	100 schools	100 schools
Extra contract teacher	100 schools		
Extra block grant	100 schools		

- Why randomize at the level of schools (not teachers?) **Much easier to implement, prevent undesirable effects within schools (spillovers, jealousy, etc.)**
- What is the difference between the individual bonus and the group bonus? **Incentive depends on own students (individual) or all students in the school (group-based)**

Balance table [Extra]

TABLE 2
SAMPLE BALANCE ACROSS TREATMENTS

	Control (1)	Group Incentive (2)	Individual Incentive (3)	p-Value (Equality of All Groups) (4)
	A. Means of Baseline Variables			
School-level variables:				
1. Total enrollment (baseline: grades 1–5)	113.2	111.3	112.6	.82
2. Total test takers (baseline: grades 2–5)	64.9	62.0	66.5	.89
3. Number of teachers	3.07	3.12	3.14	.58
4. Pupil-teacher ratio	39.5	40.6	37.5	.66
5. Infrastructure index (0–6)	3.19	3.14	3.26	.81
6. Proximity to facilities index (8–24)	14.65	14.66	14.72	.98
Baseline test performance:				
7. Math (raw %)	18.5	18.0	17.5	.69
8. Math (normalized; in SD)	.032	.001	−.032	.70
9. Telugu (raw %)	35.1	34.9	33.5	.52
10. Telugu (normalized; in SD)	.026	.021	−.046	.53

- Does this remind you of the problem set?
- If we test 100 baseline characteristics: How often would we expect p-values below 0.05 if our randomization worked?
- Do you think randomization worked in this case?

Regression specification (group work) [SA7-Q2c]

$$T_{ijkm}(Y_n) = \alpha + \gamma \cdot T_{ijkm}(Y_0) + \delta \cdot \text{Incentives} + \beta \cdot Z_m + \varepsilon_k + \varepsilon_{jk} + \varepsilon_{ijk} \quad (1)$$

The main dependent variable of interest is T_{ijkm}, which is the normalized test score on the specific subject, where i, j, k, and m denote the student, grade, school, and mandal, respectively. The term Y_0 indicates the baseline tests, and Y_n indicates a test at the end of n years of the program. Including the normalized baseline test score improves efficiency as a result of the autocorrelation between test scores across multiple periods.[21] All regressions include a set of mandal-level dummies (Z_m), and the standard errors are clustered at the school level. We also run the regressions with and without controls for household and school variables. The Incentives variable is a dummy at the school level indicating treatment status, and the parameter of interest is δ, which is the effect on test scores of being in an incentive school. The random assignment of the incentive program ensures that this is an unbiased and consistent estimate of the 1-year and 2-year treatment effects.

Regression specification (group work) [SA7-Q2c]

$$Y_{ijsm}(1) = \alpha + \gamma \cdot Y_{ijsm}(0) + \delta \cdot \text{Incentives} + \beta \cdot Z_s + \varepsilon_s + \varepsilon_j + \varepsilon_{ijs} \quad (1)$$

The main dependent variable of interest is Y_{ijsm}, which is the normalized test score on the specific subject, where i, j, s, and m denote the student, grade, school, and mandal, respectively. The term Y_0 indicates the baseline test, and Y_1 indicates a test at the end of 1 years of the program. Including the normalized baseline test score improves efficiency as a result of the autocorrelation between test scores across multiple periods.[14] All regressions include a set of mandal-level dummies (Z_s), and the standard errors are clustered at the school level. We also run the regressions with and without controls for household and school variables. The Incentives variable is a dummy at the school level indicating treatment status, and the parameter of interest is δ, which is the effect on test scores of being in an incentive school. The random assignment of the incentive program ensures that this is an unbiased and consistent estimate of the 1-year and 2-year treatment effects.

For each of the four terms (yellow, red, blue, green), think:

- How is it measured?
- Why is it included?
- What does the coefficient on the term tell us?

Regression specification (group work) [SA7-Q2c]

$$T_{ijkm}(Y_n) = \alpha + \gamma \cdot T_{ijkm}(Y_0) + \delta \cdot \text{Incentives} + \beta \cdot Z_m + \varepsilon_k + \varepsilon_{jk} + \varepsilon_{ijk} \quad (1)$$

The main dependent variable of interest is T_{ijkm}, which is the normalized test score on the specific subject, where i, j, k, and m denote the student, grade, school, and mandal, respectively. The term Y_0 indicates the baseline tests, and Y_n indicates a test at the end of n years of the program. Including the normalized baseline test score improves efficiency as a result of the autocorrelation between test scores across multiple periods.[21] All regressions include a set of mandal-level dummies (Z_m), and the standard errors are clustered at the school level. We also run the regressions with and without controls for household and school variables. The Incentives variable is a dummy at the school level indicating treatment status, and the parameter of interest is δ, which is the effect on test scores of being in an incentive school. The random assignment of the incentive program ensures that this is an unbiased and consistent estimate of the 1-year and 2-year treatment effects.

Regression specification [SA7-Q2c]

$$T_{ijkm}(Y_n) = \alpha + \gamma \cdot T_{ijkm}(Y_0) + \delta \cdot \text{ Incentives } + \beta \cdot Z_m + \varepsilon_k + \varepsilon_{jk} + \varepsilon_{ijk}.$$

- Outcome variable: $T_{ijkm}(Y_n)$: The test score of student i in grade k in school k in mandal (subdistrict) m, at the end of n years. **Normalized**.
- $T_{ijkm}(Y_0)$: test score at the *baseline* survey, for efficiency.
- *Incentives* is the variable of interest: an indicator (dummy) equal to 1 if a school was in the treatment group (and their teachers got incentive pay), and 0 otherwise. δ tells us the causal effect of performance pay for teachers on the test score of their students.
- Z_m is a set of fixed effects (dummy variables) for the Mandal, included for efficiency.
- All the ϵ terms are error terms.

Regression table [SA7-Q2d]

TABLE 3
IMPACT OF INCENTIVES ON STUDENT TEST SCORES
Dependent Variable: Normalized End-of-Year Test Score

	YEAR 1 ON YEAR 0		YEAR 2 ON YEAR 0	
	(1)	(2)	(3)	(4)
	A. Combined (Math and Language)			
Normalized lagged test score	.503***	.498***	.452***	.446***
	(.013)	(.013)	(.015)	(.015)
Incentive school	.149***	.165***	.219***	.224***
	(.042)	(.042)	(.047)	(.048)
School and household controls	No	Yes	No	Yes
Observations	42,145	37,617	29,760	24,665
R^2	.31	.34	.24	.28

- What are the rows and columns?
- Coefficient on normalized test score, significance?
- What is the coefficient on incentive school? Significant?
- Where do we see control variables in the table?
  How do they affect our coefficient of interest (δ)?

Regression table [SA7-Q2d]

- What are the rows and columns?

Regression table [SA7-Q2d]

- What are the rows and columns? The four columns are four different regression specifications (with and without controls, after one year and after two years). The rows give coefficients and their standard errors, and some additional information.
- Coefficient on normalized test score, significance?

Regression table [SA7-Q2d]

- What are the rows and columns? The four columns are four different regression specifications (with and without controls, after one year and after two years). The rows give coefficients and their standard errors, and some additional information.
- Coefficient on normalized test score, significance? The coefficient is around 0.5, with a standard error of 0.013. Significant at the 1% level. The coefficient means that students with a one standard deviation higher test score at the baseline are expected to have a 0.5 standard deviation higher test score at the endline, on average. Test scores are persistent, but correlation of test scores over time is not perfect.

Regression table [SA7-Q2d]

- What is the coefficient on incentive school? Significant?

Regression table [SA7-Q2d]

- What is the coefficient on incentive school? Significant?

  Around 0.15 without additional control variables, and 0.165 with control variables in year 1. The bonus incentive treatment has led to an increase in test scores by around 0.15 (0.165) standard deviations - a sizeable effect! In year 2, the treatment effect is even bigger (0.22 sd). The coefficient is always significant at the 1% level.

Regression table [SA7-Q2d]

- What is the coefficient on incentive school? Significant?
  Around 0.15 without additional control variables, and 0.165 with control variables in year 1. The bonus incentive treatment has led to an increase in test scores by around 0.15 (0.165) standard deviations - a sizeable effect! In year 2, the treatment effect is even bigger (0.22 sd). The coefficient is always significant at the 1% level.
- Where do we see control variables in the table?
  How do they affect our coefficient of interest (δ)?

Regression table [SA7-Q2d]

- What is the coefficient on incentive school? Significant?

  Around 0.15 without additional control variables, and 0.165 with control variables in year 1. The bonus incentive treatment has led to an increase in test scores by around 0.15 (0.165) standard deviations - a sizeable effect! In year 2, the treatment effect is even bigger (0.22 sd). The coefficient is always significant at the 1% level.
- Where do we see control variables in the table? How do they affect our coefficient of interest (δ)?

  Control variables are omitted from the table output, but we see that they slightly change the coefficient. The coefficient on the treatment becomes larger, but not by much. If including them changed a lot, we would question our assumptions about the RCT!

Benchmarking the effect [SA7-Q2e]

TABLE 10
IMPACT OF INPUTS VERSUS INCENTIVES ON LEARNING OUTCOMES
Dependent Variable: Normalized End-of-Year Test Score

	Year 1 on Year 0			Year 2 on Year 0		
	Combined (1)	Math (2)	Language (3)	Combined (4)	Math (5)	Language (6)
Normalized lagged score	.512***	.494***	.536***	.458***	.416***	.499***
	(.010)	(.012)	(.011)	(.012)	(.016)	(.012)
Incentives	.15***	.179***	.121***	.218***	.272***	.164***
	(.041)	(.048)	(.039)	(.049)	(.057)	(.046)
Inputs	.102***	.117***	.086**	.085*	.089*	.08*
	(.038)	(.042)	(.037)	(.046)	(.052)	(.044)
F-statistic *p*-value (inputs = incentives)	.178	.135	.298	.003	.000	.044
Observations	69,157	34,376	34,781	49,503	24,628	24,875
R^2	.30	.29	.32	.225	.226	.239

NOTE.—These regressions pool data from all 500 schools in the study: group and individual incentive treatments are pooled together as incentives, and the extra contract teacher and block grant treatments are pooled together as inputs. All regressions include mandal (subdistrict) fixed effects and standard errors clustered at the school level.

Benchmarking the effect [SA7-Q2e]

TABLE 10
IMPACT OF INPUTS VERSUS INCENTIVES ON LEARNING OUTCOMES
Dependent Variable: Normalized End-of-Year Test Score

	Year 1 on Year 0			Year 2 on Year 0		
	Combined (1)	Math (2)	Language (3)	Combined (4)	Math (5)	Language (6)
Normalized lagged score	.512***	.494***	.536***	.458***	.416***	.499***
	(.010)	(.012)	(.011)	(.012)	(.016)	(.012)
Incentives	.13***	.179***	.121***	.218***	.272***	.164***
	(.011)	(.048)	(.039)	(.049)	(.057)	(.046)
Inputs	.102***	.117***	.086**	.085*	.089*	.08*
	(.038)	(.042)	(.037)	(.046)	(.052)	(.044)
F-statistic *p*-value (inputs = incentives)	.178	.135	.298	.003	.000	.044
Observations	69,157	34,376	34,781	49,503	24,628	24,875
R^2	.30	.29	.32	.225	.226	.239

NOTE.—These regressions pool data from all 500 schools in the study: group and individual incentive treatments are pooled together as incentives, and the extra contract teacher and block grant treatments are pooled together as inputs. All regressions include mandal (subdistrict) fixed effects and standard errors clustered at the school level.

- What was the effect of the additional inputs?
- Which program did better? By how much?
- Can you tell whether the difference is significant?

Benchmarking the effect [SA7-Q2e]

- What was the effect of the additional inputs? Additional inputs increased test scores by around 0.1 standard deviations in the first year (a bit more for maths than for language). This is significant at the 1% level. Compared to year 0, test scores in year 2 were around 0.09 standard deviations higher, significant at the 10%-level.
- Which program did better? By how much? Comparing the two coefficients, incentives did better than inputs throughout. The difference is around 0.05 sd in the first year and 0.14 in the second year.
- Can you tell whether the difference is significant? The row "*F*-statistic *p*-value" gives the p-value for a statistical test whether the two coefficients are the same. We fail to reject this null hypothesis in year 1, but reject it at the 5% level in year 2.

Conclusion

What do we learn?

- We evaluated an RCT that is representative of schools in an Indian state with 80 million population
- The randomization design allows us to compare various treatment arms against each other
- The non-significant differences in the balance table indicate that randomization was successful
- We found sizeable effects of the pay-for-performance incentive scheme: The students in treated schools did significantly better on standardized tests
- The effect is larger than giving a comparable amount of money to the school directly
- **This study provides causal evidence that pay-for-performance for teachers increased students' outcomes**

Comments and criticisms?

- External validity: Would this also apply in other Indian states? Other developing countries? The world in general?
- Is this feasible in settings with no standardized tests?
- How do the costs compare against the benefits?
- . . .

Other tables if there is extra time

[Extra] Individual vs. group incentive effects

TABLE 8
GROUP VERSUS INDIVIDUAL INCENTIVES
Dependent Variable: Normalized End-of-Year Test Score

	YEAR 1 ON YEAR 0			YEAR 2 ON YEAR 0		
	Combined (1)	Math (2)	Telugu (3)	Combined (4)	Math (5)	Telugu (6)
Individual incentive school	.156***	.184***	.130***	.283***	.329***	.239***
	(.050)	(.059)	(.045)	(.058)	(.067)	(.054)
Group incentive school	.141***	.175***	.107**	.154***	.216***	.092*
	(.050)	(.057)	(.047)	(.057)	(.068)	(.052)
F-statistic *p*-value (testing group incentive school = individual incentive school)	.765	.889	.610	.057	.160	.016
Observations	42,145	20,946	21,199	29,760	14,797	14,963
R^2	.31	.299	.332	.25	.25	.26

NOTE.—All regressions include mandal (subdistrict) fixed effects and standard errors clustered at the school level.

[Extra] Teaching to the test: Repeat vs. non-repeat

TABLE 4
IMPACT OF INCENTIVES BY REPEAT AND NONREPEAT QUESTIONS
Dependent Variable: Percentage Score

	COMBINED		MATH		TELUGU	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
Percentage score on non-repeat questions	.335*** (.007)	.328*** (.007)	.256*** (.007)	.257*** (.008)	.414*** (.008)	.397*** (.007)
Percentage score on repeat questions	.352*** (.006)	.42*** (.005)	.252*** (.007)	.386*** (.006)	.452*** (.007)	.468*** (.007)
Incremental score in incentive schools for non-repeats	.030*** (.009)	.039*** (.009)	.033*** (.009)	.046*** (.010)	.027*** (.010)	.033*** (.010)
Incremental score in incentive schools for repeats	.043*** (.011)	.043*** (.011)	.042*** (.013)	.044*** (.012)	.043*** (.011)	.041*** (.013)
Test for equality of treatment effect for repeat and nonrepeat questions (F-statistic, p-value)	.141	.584	.374	.766	.076	.354
Observations	62,872	54,972	31,225	29,594	31,647	25,378
R^2	.24	.18	.26	.23	.29	.18

NOTE.—Repeat questions are questions that at the time of administering the particular test had appeared identically on any earlier test (across grades).

[Extra] Heterogeneous treatment effects A

TABLE 6
HETEROGENEOUS TREATMENT EFFECTS
A. HOUSEHOLD AND SCHOOL CHARACTERISTICS

	Log School Enrollment (1)	School Proximity (8–21) (2)	School Infrastructure (0–6) (3)	Household Affluence (0–7) (4)	Parental Literacy (0–1) (5)	Scheduled Caste/Tribe (6)	Male (7)	Normalized Baseline Score (8)
				Two-Year Effect				
Incentive	−.198	−.019	.28**	.09	.221***	.296***	.335***	.219***
	(.354)	(.199)	(.130)	(.073)	(.051)	(.039)	(.049)	(.047)
Covariate	−.065	−.005	.025	.017	.068***	−.066	.029	.448***
	(.058)	(.010)	(.038)	(.014)	(.015)	(.042)	(.027)	(.024)
Interaction	.083	.018	−.02	.058**	−.003	−.013	−.02	.006
	(.074)	(.014)	(.010)	(.019)	(.019)	(.056)	(.034)	(.031)
Observations	29,760	29,760	29,760	25,231	25,226	29,760	25,881	29,760
R^2	.244	.244	.243	.272	.273	.244	.266	.243
				One-Year Effect				
Incentive	−.36	−.076	.032	.001	.166***	.164***	.157***	.149***
	(.381)	(.161)	(.110)	(.060)	(.017)	(.015)	(.044)	(.042)
Covariate	−.128**	−.016*	−.001	.017	.08***	.007	.016	.502***
	(.061)	(.008)	(.025)	(.013)	(.012)	(.035)	(.020)	(.021)
Interaction	.103	.017	.011	.042**	−.018	−.06	.002	.000
	(.081)	(.011)	(.031)	(.017)	(.016)	(.048)	(.025)	(.026)
Observations	42,145	41,131	41,131	38,345	38,325	42,145	39,540	42,145
R^2	.31	.32	.32	.34	.34	.31	.33	.31

[Extra] Heterogeneous treatment effects B

B. Teacher Characteristics (Pooled Regression Using Both Years of Data)

	Education (1)	Training (2)	Years of Experience (3)	Salary (Log) (4)	Male (5)	Teacher Absence (6)	Active Teaching (7)	Active or Passive Teaching (8)
Incentive	−.118	−.224	.258***	1.775**	.031	.15***	.084	.118
	(.163)	(.176)	(.059)	(.828)	(.091)	(.050)	(.054)	(.074)
Covariate	.003	.051	.001	.051	.084	.119	.055	.151
	(.052)	(.041)	(.003)	(.066)	(.057)	(.187)	(.078)	(.098)
Interaction	.086*	.138**	−.009**	−.179*	.09	.013	.164*	.064
	(.050)	(.061)	(.004)	(.091)	(.069)	(.171)	(.098)	(.111)
Observations	53,757	53,890	54,142	53,122	54,142	53,609	53,385	53,383
R^2	.29	.29	.29	.29	.29	.29	.29	.29

Introduction to Instrumental Variables

Econ 140, Section 7

Jonathan Old

Roadmap

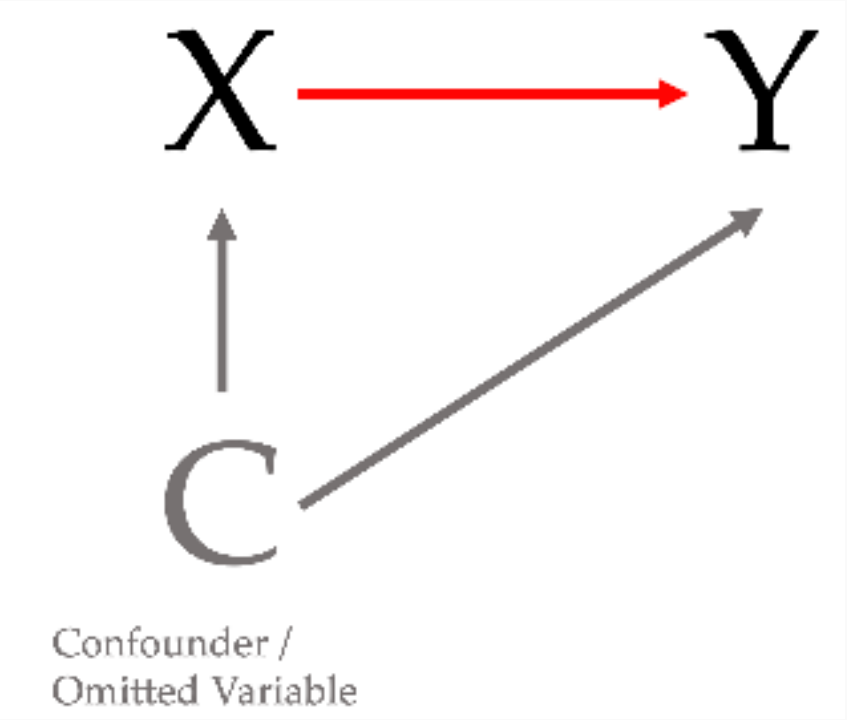
1. Introduction to Instrumental Variables
2. IV Conditions
3. IV Summary
4. Group work
5. Section Assignment

Any questions?

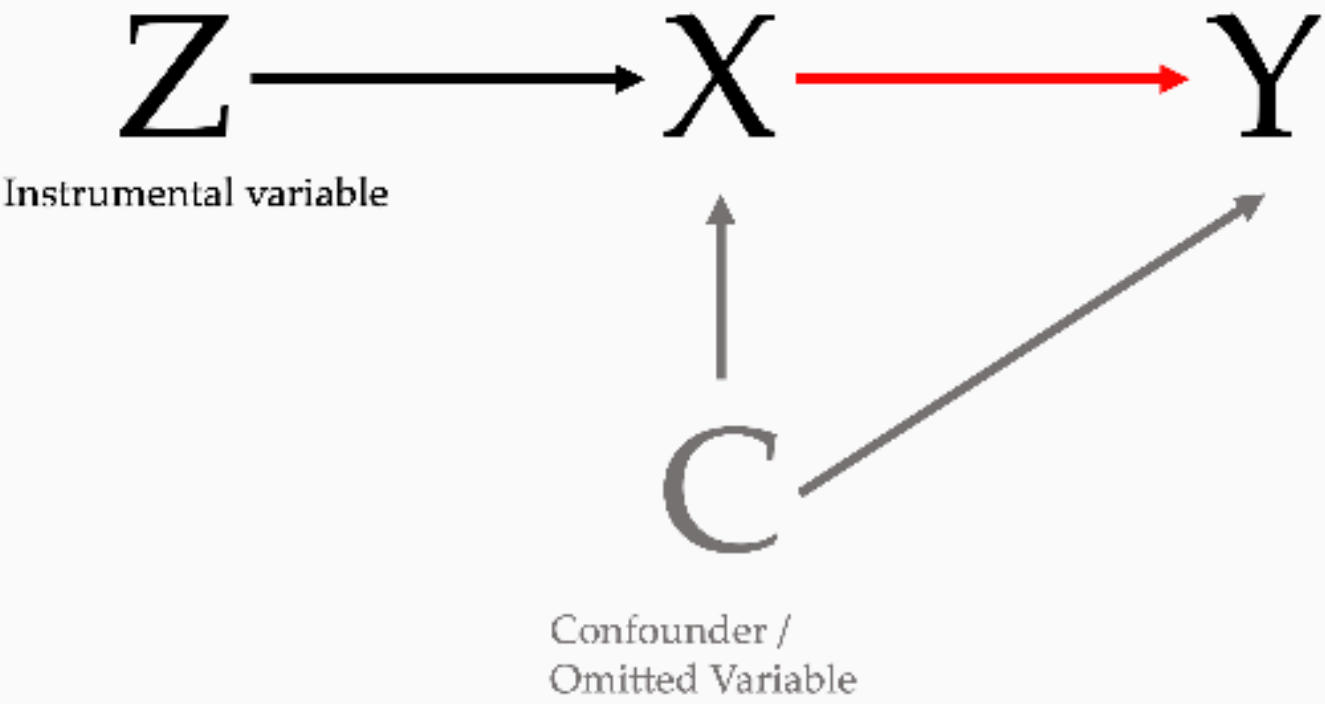
… Remember – Every question is useful!

Introduction to Instrumental Variables

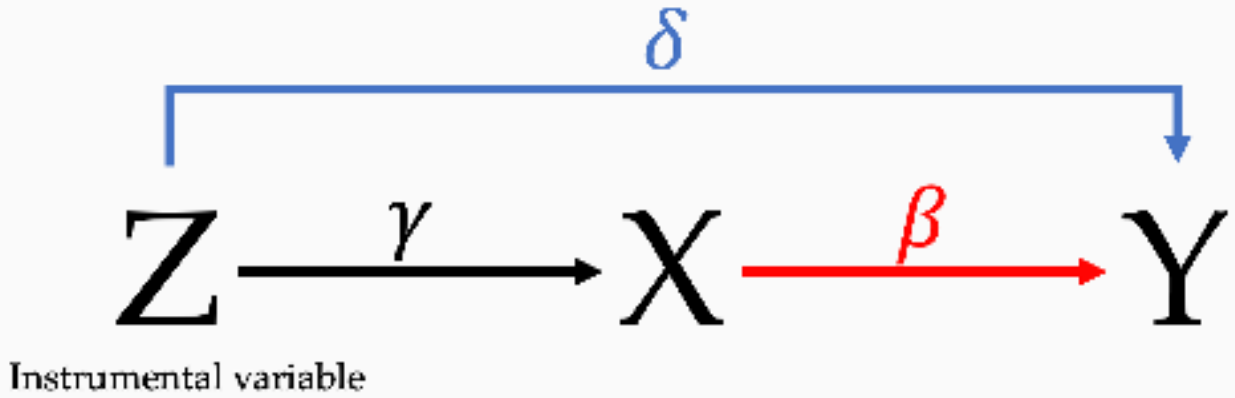
Recap: Omitted Variable Bias



Instrumental variables: The setup



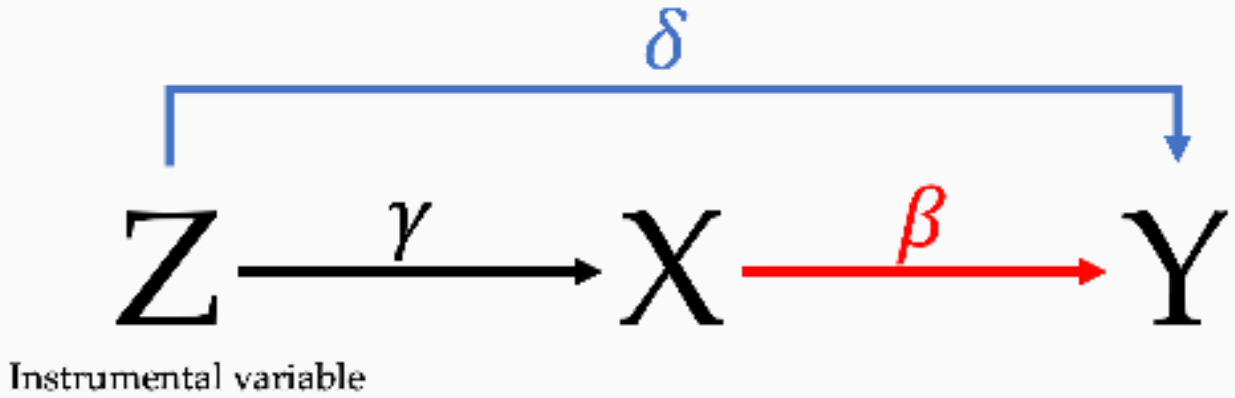
Recap: IV ”rescales” the effect



A simple example:

- We want to know the effect of chocolate (X) on happiness (Y), using a randomized voucher as instrument (Z).
- We find: people with voucher were 3 points more happy ($\delta = 3$), and ate 0.5 more chocolates ($\gamma = 0.5$).
- Then, the effect of eating one more chocolate is:

Recap: IV ”rescales” the effect



A simple example:

- We want to know the effect of chocolate (X) on happiness (Y), using a randomized voucher as instrument (Z).
- We find: people with voucher were 3 points more happy ($\delta = 3$), and ate 0.5 more chocolates ($\gamma = 0.5$).
- Then, the effect of eating one more chocolate is: $\beta = \delta/\gamma = 3/0.5 = 6$.

Calculating the IV coefficient

What is the effect of **eating chocolate** (D) on happiness (Y).

- Why not estimate: $Y_i = \alpha + \beta D_i + \varepsilon_i$?

Randomly give voucher to buy chocolate at 90% discount (Z).

- Why not estimate: $Y_i = \alpha + \beta Z_i + \varepsilon_i$?

Let us set up some regressions:

$$\text{Regression of interest: } Y_i = \alpha + \beta D_i + e_i$$

$$\text{First stage: } D_i = \alpha_1 + \gamma Z_i + u_i$$

$$\text{Reduced Form: } Y_i = \alpha_2 + \delta Z_i + v_i$$

$$\text{Plug in regression of interest: } Y_i = \alpha + \beta(\alpha_1 + \gamma \cdot Z_i + u_i) + e_i$$

$$\text{Get back reduced form: } = \underbrace{(\alpha + \beta\alpha_1)}_{\alpha_2} + \underbrace{(\beta\gamma)}_{\delta} Z_i + \underbrace{(\beta u_i + e_i)}_{v_i}$$

$$\text{So we see that } \quad \delta = \beta\gamma \Leftrightarrow \beta = \delta/\gamma$$

Interpretation of the IV coefficient

- How do we interpret γ ?

Interpretation of the IV coefficient

- How do we interpret γ ? The average difference in chocolate consumption between those who got a voucher and those who didn't

Interpretation of the IV coefficient

- How do we interpret γ ? The average difference in chocolate consumption between those who got a voucher and those who didn't
- How do we interpret δ ?

Interpretation of the IV coefficient

- How do we interpret γ ? The average difference in chocolate consumption between those who got a voucher and those who didn't
- How do we interpret δ ? The average difference in happiness between those who got a voucher and those who didn't

Interpretation of the IV coefficient

- How do we interpret γ ? The average difference in chocolate consumption between those who got a voucher and those who didn't
- How do we interpret δ ? The average difference in happiness between those who got a voucher and those who didn't

$$\beta = \frac{\gamma}{\delta} = \frac{E\left[Y_i \mid Z_i = 1\right] - E\left[Y_i \mid Z_i = 0\right]}{E\left[D_i \mid Z_i = 1\right] - E\left[D_i \mid Z_i = 0\right]}$$

IV gives us the treatment effect for the compliers

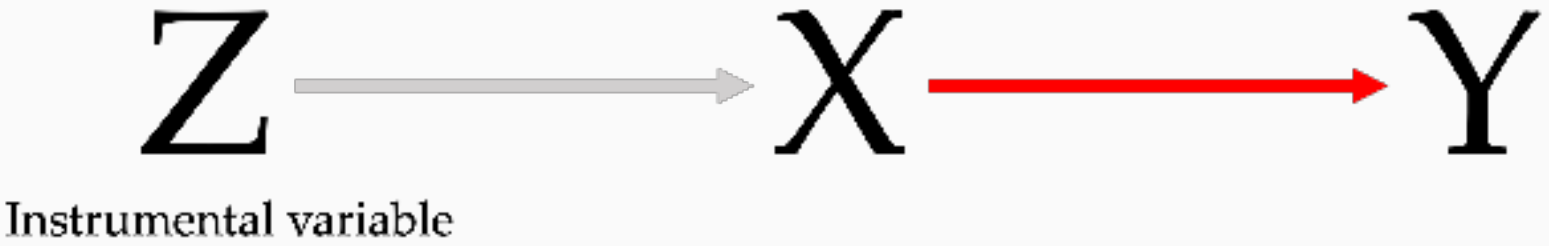
Potential outcomes! (unobserved)		***Does not get voucher (Z=0)***	
Gets voucher (Z=1)		*Eats chocolate (D=1)*	*Does not eat chocolate (D=0)*
	Eats chocolate (D=1)	Always-takers: E(D|Z=1)=E(D|Z=0)=1 ➔ E(Y|Z=1)=E(Y|Z=0)	Compliers
	Does not eat chocolate (D=0)	Defiers	Never-takers: E(D|Z=1)=E(D|Z=0)=0 ➔ E(Y|Z=1)=E(Y|Z=0)

IV Conditions

Does IV always work?

- No! It only works if we have a valid instrument
- For this, we need three conditions:

1. Relevance: *Z* must truly affect *X*



Does IV always work?

- No! It only works if we have a valid instrument
- For this, we need three conditions:

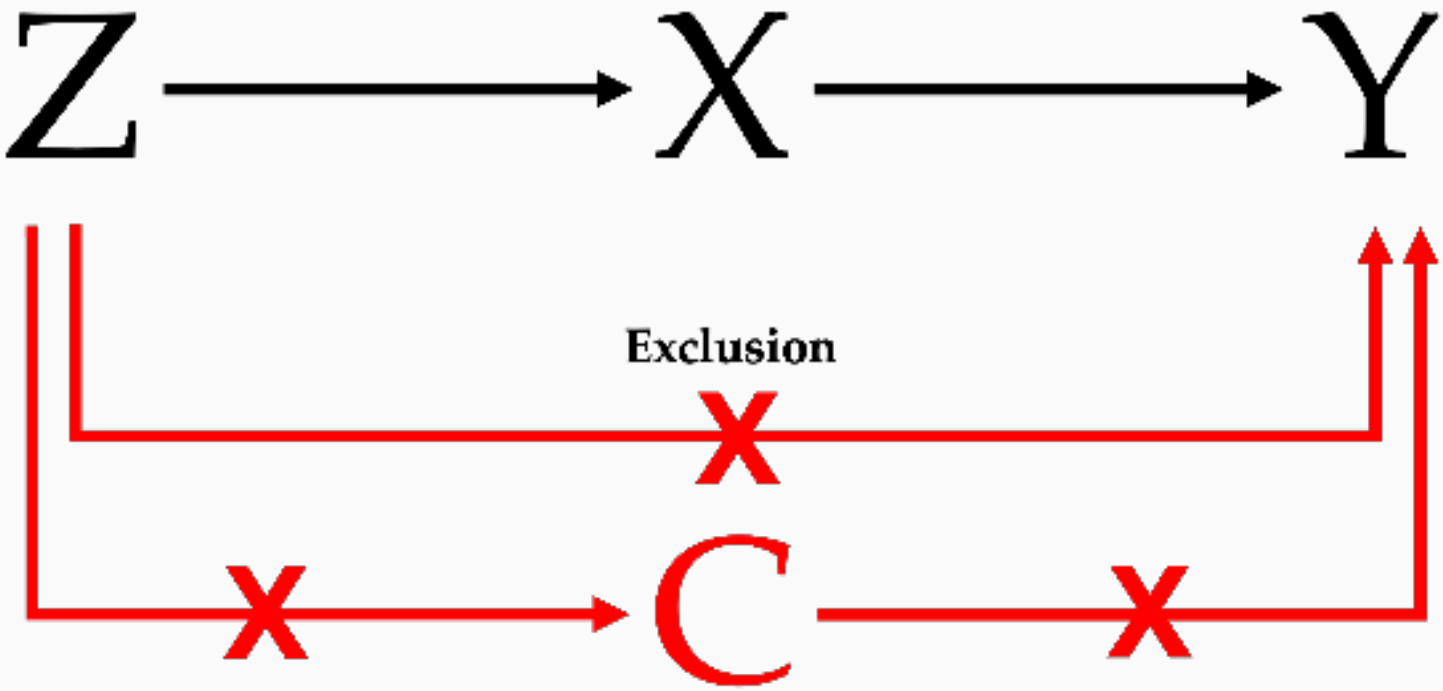
2. Independence/Exogeneity: Z is as good as randomly assigned



Does IV always work?

- No! It only works if we have a valid instrument
- For this, we need three conditions:

3. Exclusion: The **ONLY** way that *Z* affects *Y* is via *X*!

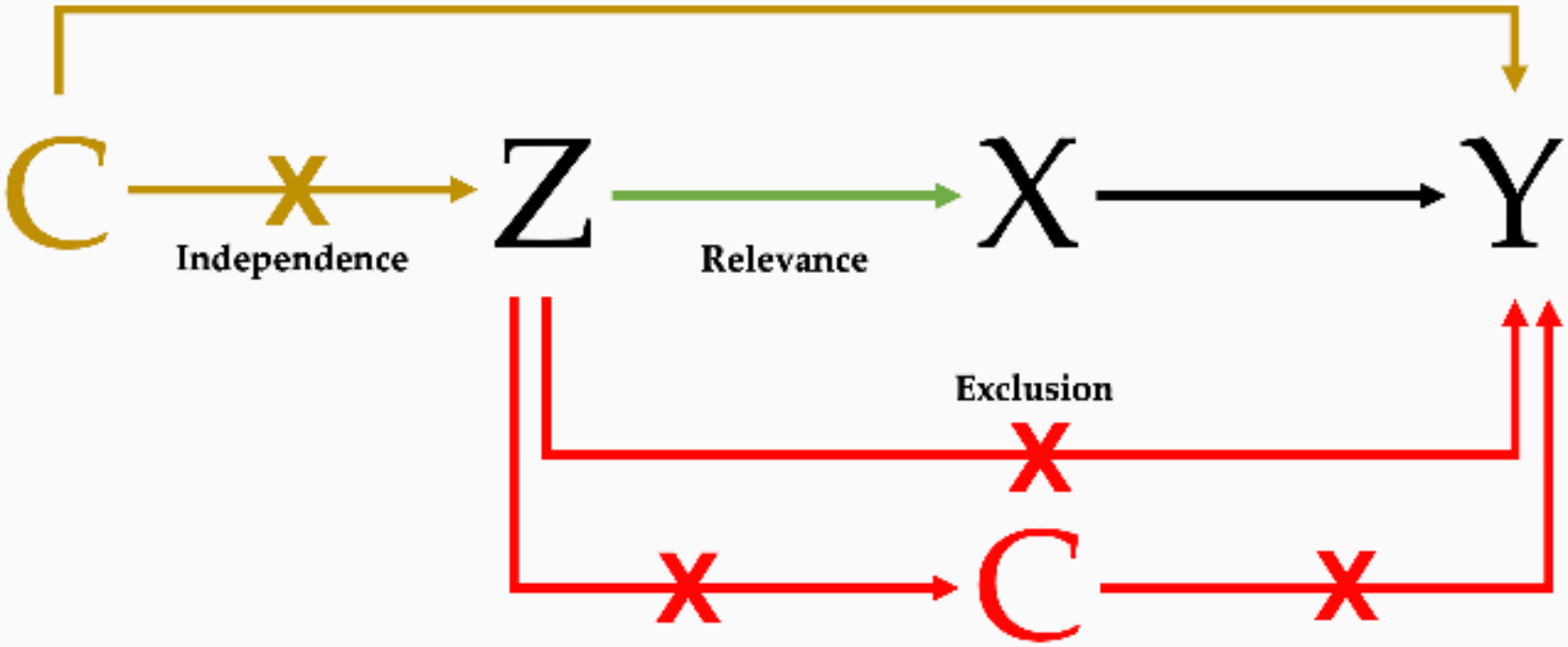


IV Summary

IV summary

We need the following three assumptions for IV to work:

1. **Relevance**: Z must truly affect X
2. **Independence/Exogeneity**: Z is as good as randomly assigned
3. **Exclusion Restriction**: The **only** way that Z affects Y is via X.



Any questions?

… Remember – Every question is useful!

Group work

Group work

Group 1: We are interested in the effect of being in the army on crime. We instrument being in the army with a lottery (paper)

Group 2: We are interested in the effect of income on conflict. We instrument income with rainfall (paper)

Group 3: We are interested in the effect of air pollution on mortality. We instrument local air pollution with wind direction (paper)

1. Relevance: Z must truly affect X
2. Independence/Exogeneity: Z is as good as randomly assigned
3. Exclusion restriction: The **only** way that Z affects Y is via X

Your job: Discuss whether these assumptions hold!

Any questions?

… Remember – Every question is useful!

Section Assignment

In R

See code in R!

Instrumental Variable: Applications

Econ 140, Section 8

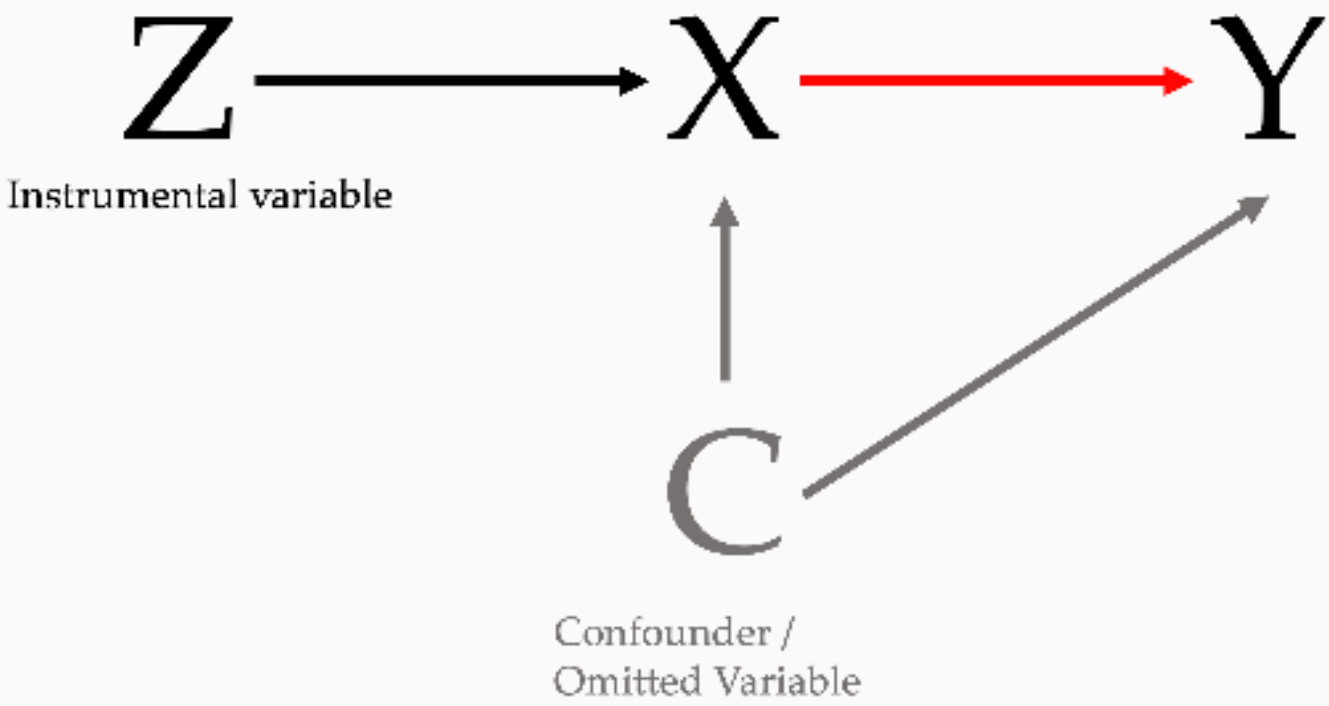
Jonathan Old

Roadmap

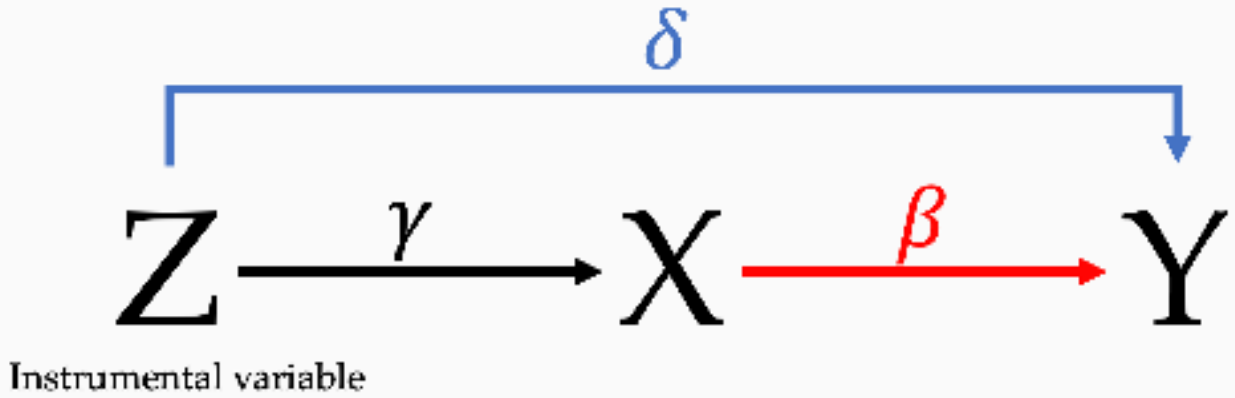
1. Recap: IV
2. Group work
3. Colonial Origins of Development [SA8-Q1]
4. LATE [SA8-Q2]

Recap: IV

Recap: IV



Recap: IV ”rescales” the effect



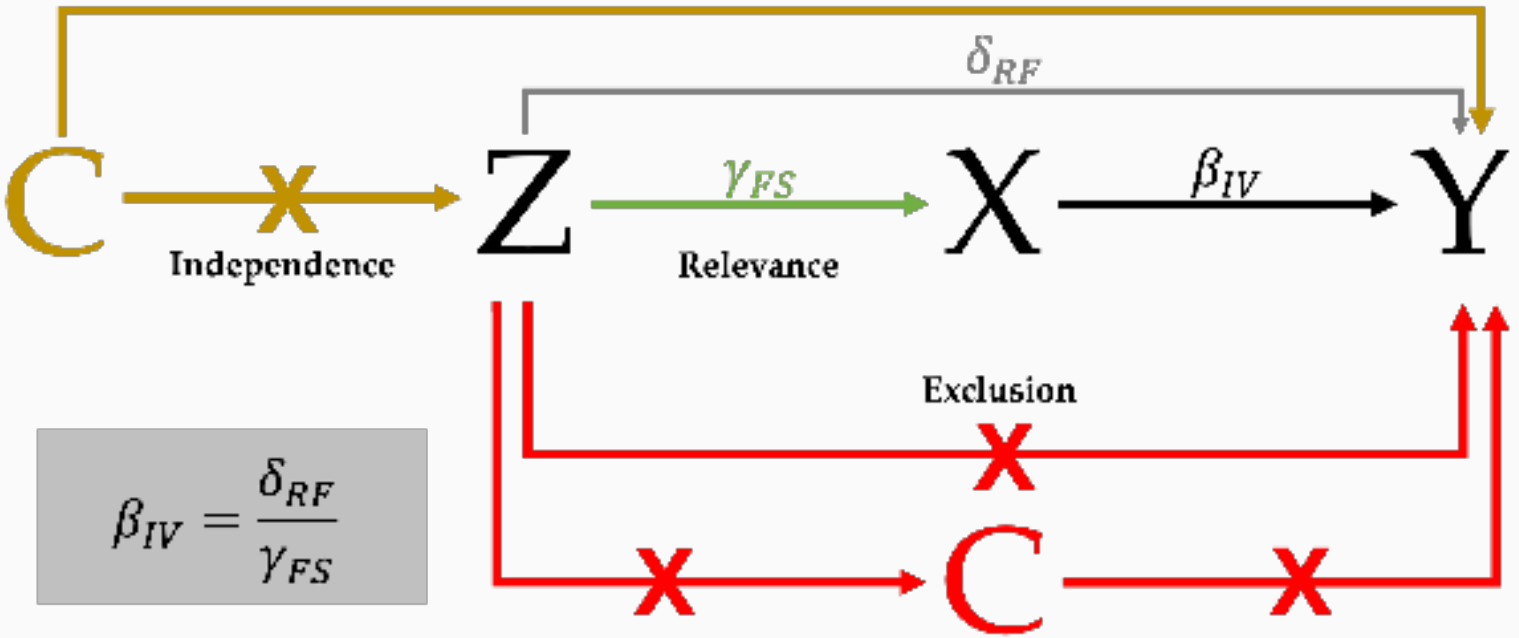
A simple example:

- We want to know the effect of chocolate (X) on happiness (Y), using a randomized voucher as instrument (Z).
- We find: people with voucher were 3 points more happy ($\delta = 3$), and ate 0.5 more chocolates ($\gamma = 0.5$).
- Then, the effect of eating one more chocolate is $\beta = \delta/\gamma = 3/0.5 = 6$.

Recap: IV summary

We need the following three assumptions for IV to work:

1. **Relevance**: Z must truly affect X
2. **Independence/Exogeneity**: Z is as good as randomly assigned
3. **Exclusion Restriction**: The **only** way that Z affects Y is via X.



$$\beta_{IV} = \frac{\delta_{RF}}{\gamma_{FS}}$$

Group work

Get into groups of 4 ...

... Your job today: Ask as many ”silly” questions as possible!

Group work

Group 1: We are interested in the effect of being in the army on crime. We instrument being in the army with a lottery (paper)

Group 2: We are interested in the effect of income on conflict. We instrument income with rainfall (paper)

Group 3: We are interested in the effect of air pollution on mortality. We instrument local air pollution with wind direction (paper)

1. Relevance: Z must truly affect X
2. Independence/Exogeneity: Z is as good as randomly assigned
3. Exclusion restriction: The **only** way that Z affects Y is via X

Your job: Discuss whether these assumptions hold!

Any questions?

… Remember: The more questions, the better!

Colonial Origins of Development [SA8-Q1]

Introduction

Settler Mortality: OLS [SA8-Q1b]

The authors are interested in the effect of good institutions (C_i) on economic development (log GDP per capita, $\ln y_i$). They also have a series of control variables X_i, and run the following regression:

$$\ln y_i = \alpha + \beta C_i + \mathbf{X_i}\gamma + u_i$$

- Interpret β.
- Suppose that institutional quality is measured from 0 (worst institutions) to 1 (best institutions). What sign do you expect β to have?
- You find that $\hat{\beta} = 0.8$. What do you conclude about the causal effect of good institutions on economic development?

Settler Mortality: OLS [SA8-Q1b]

- β measures the association between institutional quality and GDP per capita in the data. An increase in institutional quality by one unit is associated with an increase in GDP per capita by $\beta 100\%$, all else equal (the last because we have control variables included).

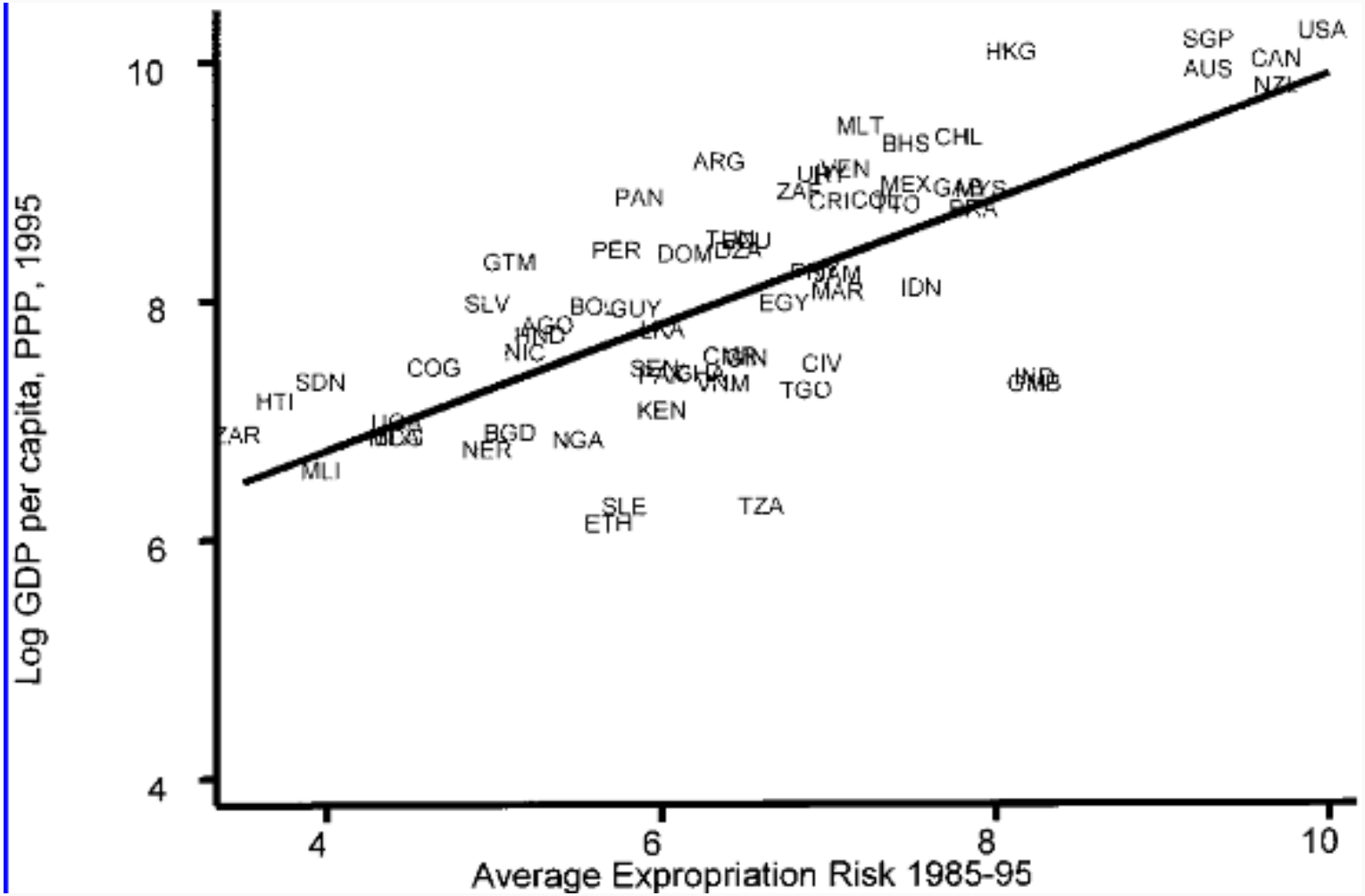
Settler Mortality: OLS [SA8-Q1b]

- β measures the association between institutional quality and GDP per capita in the data. An increase in institutional quality by one unit is associated with an increase in GDP per capita by $\beta 100\%$, all else equal (the last because we have control variables included).
- In the world, we typically see that richer countries have better institutions. Therefore, we expect β to be positive.

Settler Mortality: OLS [SA8-Q1b]

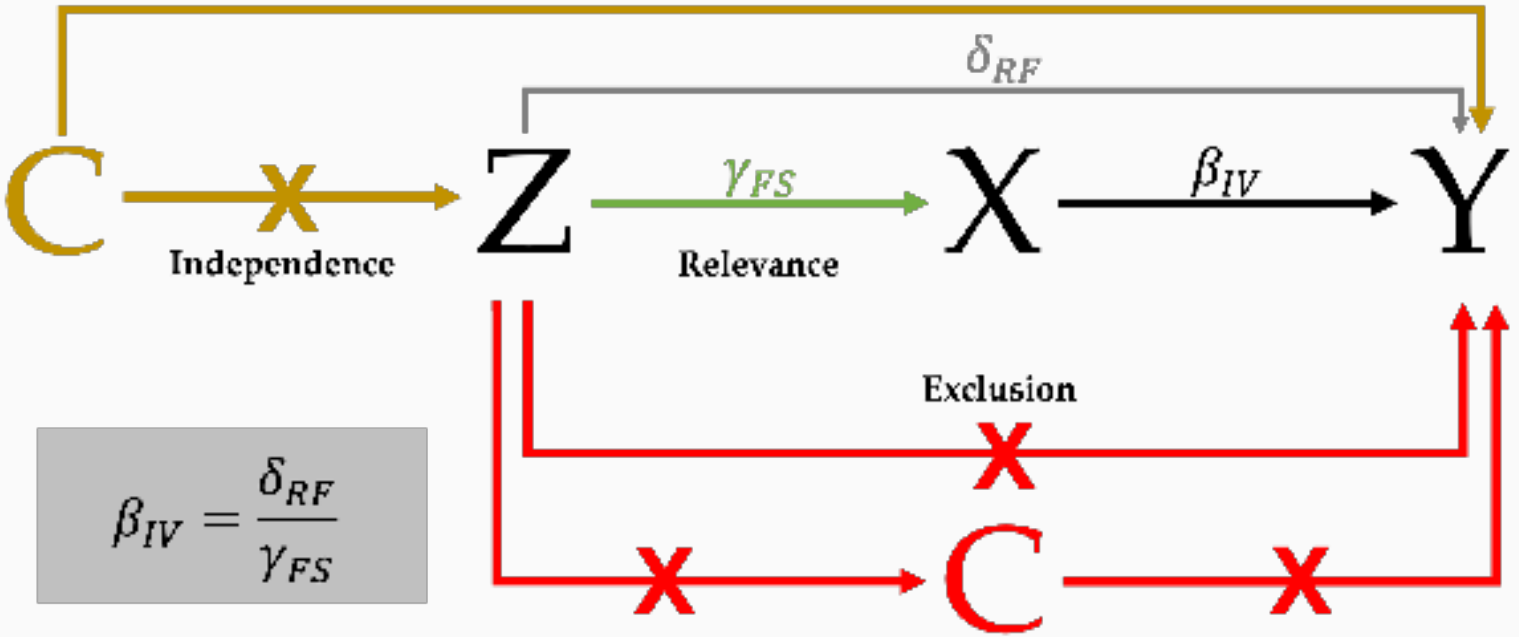
- β measures the association between institutional quality and GDP per capita in the data. An increase in institutional quality by one unit is associated with an increase in GDP per capita by $\beta 100\%$, all else equal (the last because we have control variables included).
- In the world, we typically see that richer countries have better institutions. Therefore, we expect β to be positive.
- The coefficient means that, in the data, we observe that the countries with the best institutions have around 80% higher GDP per capita than the countries with the worst institutions. However, since there are omitted variables, this does not tell us anything about the causal effect of institutions.

Settler Mortality: OLS [SA8-Q1b]



Settler Mortality: IV Approach [SA8-Q1c]

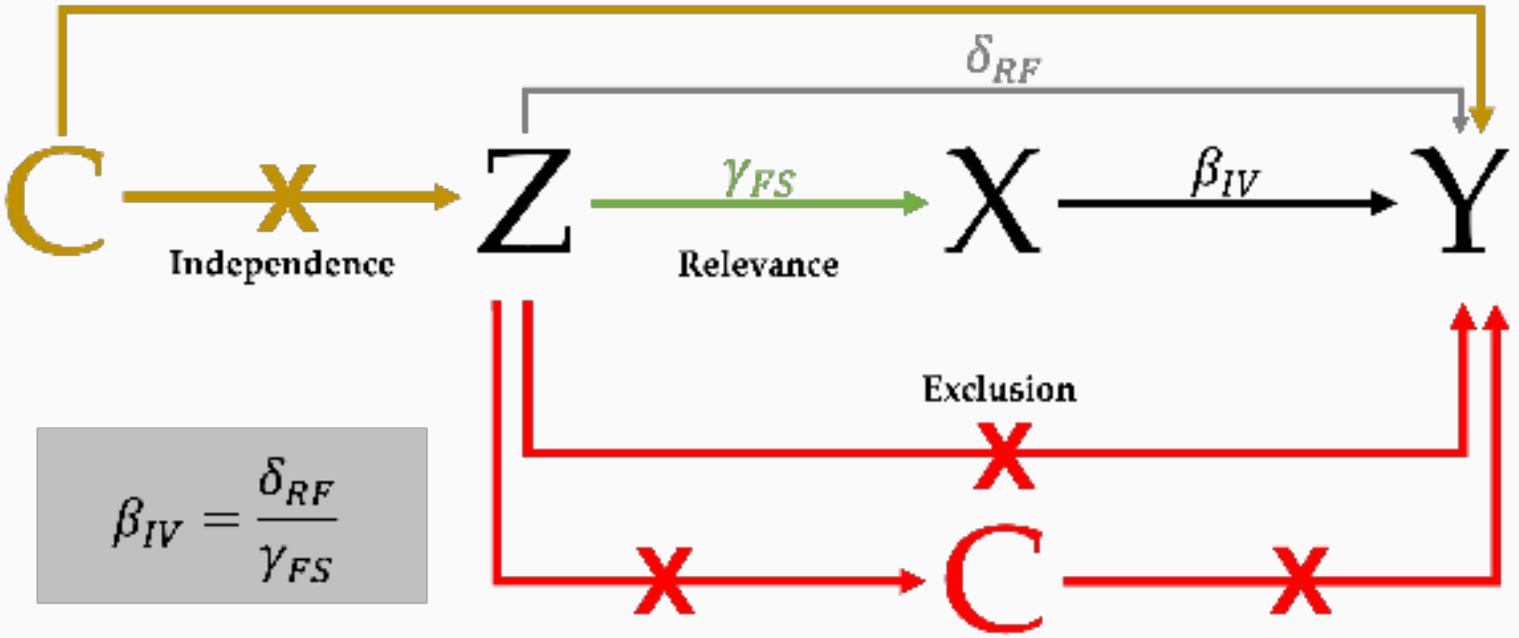
- The authors propose to use an instrumental variables approach. They use settler mortality (S_i) as an instrument for the quality of institutions (C_i).
- (In Groups:) Describe in words the three IV assumptions that settler mortality has to fulfil here to be valid.



$$\beta_{IV} = \frac{\delta_{RF}}{\gamma_{FS}}$$

Settler Mortality: IV Approach [SA8-Q1d]

- (In Groups:) For each of the three IV conditions, give one argument why they hold, and one example how they could be violated



$$\beta_{IV} = \frac{\delta_{RF}}{\gamma_{FS}}$$

Q1c,d: Sketch of solutions

1. Relevance: historical settler mortality and current institutions must be strongly related (an empirical question we can verify. Maybe, conditional on control variables, they are not!)

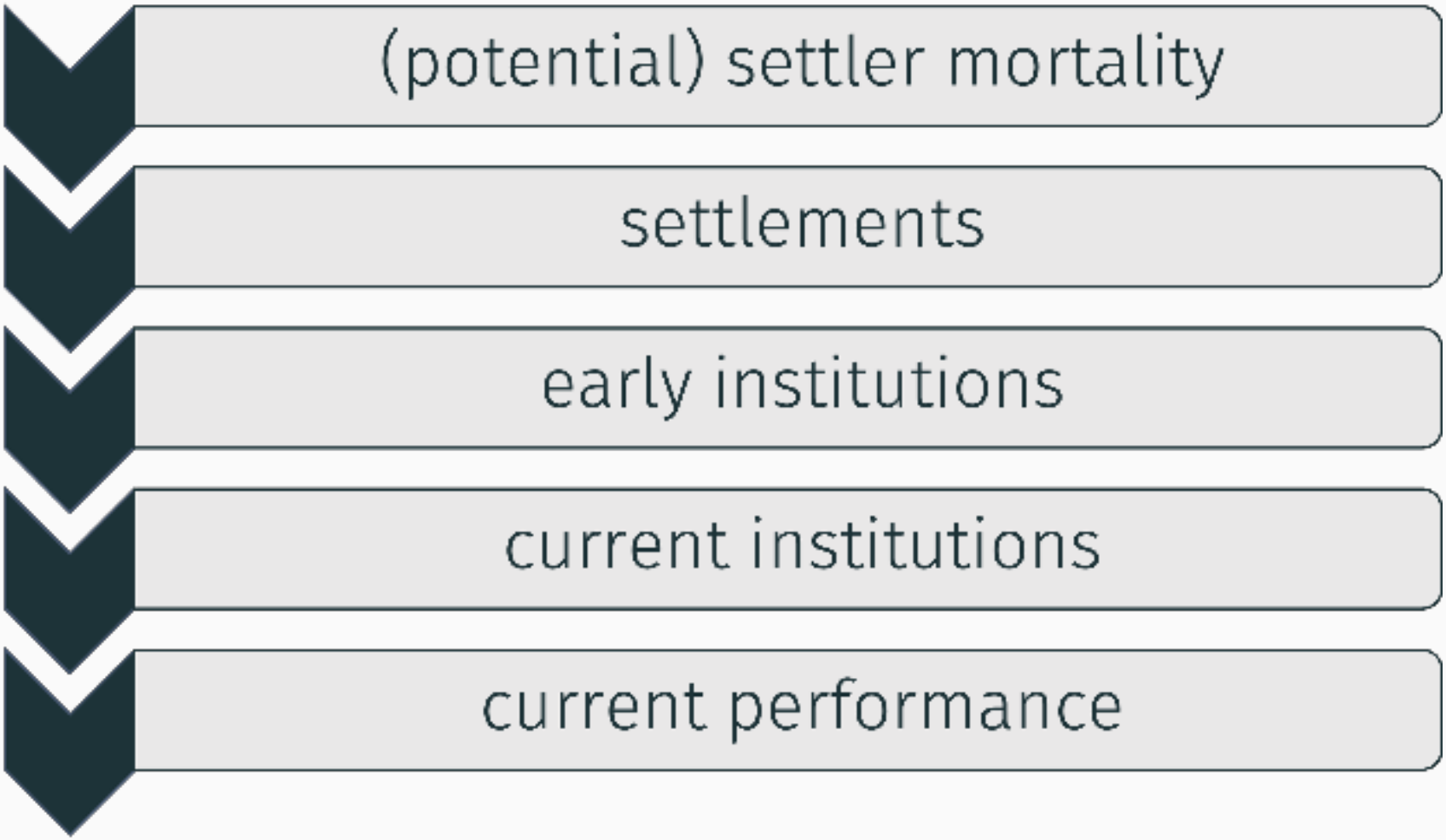
Q1c,d: Sketch of solutions

1. Relevance: historical settler mortality and current institutions must be strongly related (an empirical question we can verify. Maybe, conditional on control variables, they are not!)
2. Independence (Exogeneity): historical settler mortality cannot be correlated with anything (besides current institutions and our control variables) that impact GDP (what if climate or temperature affects both settler mortality and GDP today – e.g., via malaria)

Q1c,d: Sketch of solutions

1. Relevance: historical settler mortality and current institutions must be strongly related (an empirical question we can verify. Maybe, conditional on control variables, they are not!)
2. Independence (Exogeneity): historical settler mortality cannot be correlated with anything (besides current institutions and our control variables) that impact GDP (what if climate or temperature affects both settler mortality and GDP today – e.g., via malaria)
3. Exclusion: historical settler mortality can only impact GDP per capita through its impact on current institutions (What if settler mortality also led to different probabilities of being colonized?)

Theory


(potential) settler mortality
settlements
early institutions
current institutions
current performance

Settler Mortality: IV Relevance [SA8-Q1e]

- How would you write out the regression equation to estimate whether the first stage is strong?

Settler Mortality: IV Relevance [SA8-Q1e]

- How would you write out the regression equation to estimate whether the first stage is strong?

$$C_i = \delta + \eta S_i + \mathbf{X_i}\theta + \nu_i$$

Settler Mortality: IV Relevance [SA8-Q1e]

- How would you write out the regression equation to estimate whether the first stage is strong?

$$C_i = \delta + \eta S_i + \mathbf{X_i}\theta + \nu_i$$

- What coefficient should we look at to evaluate the strength of the first stage?

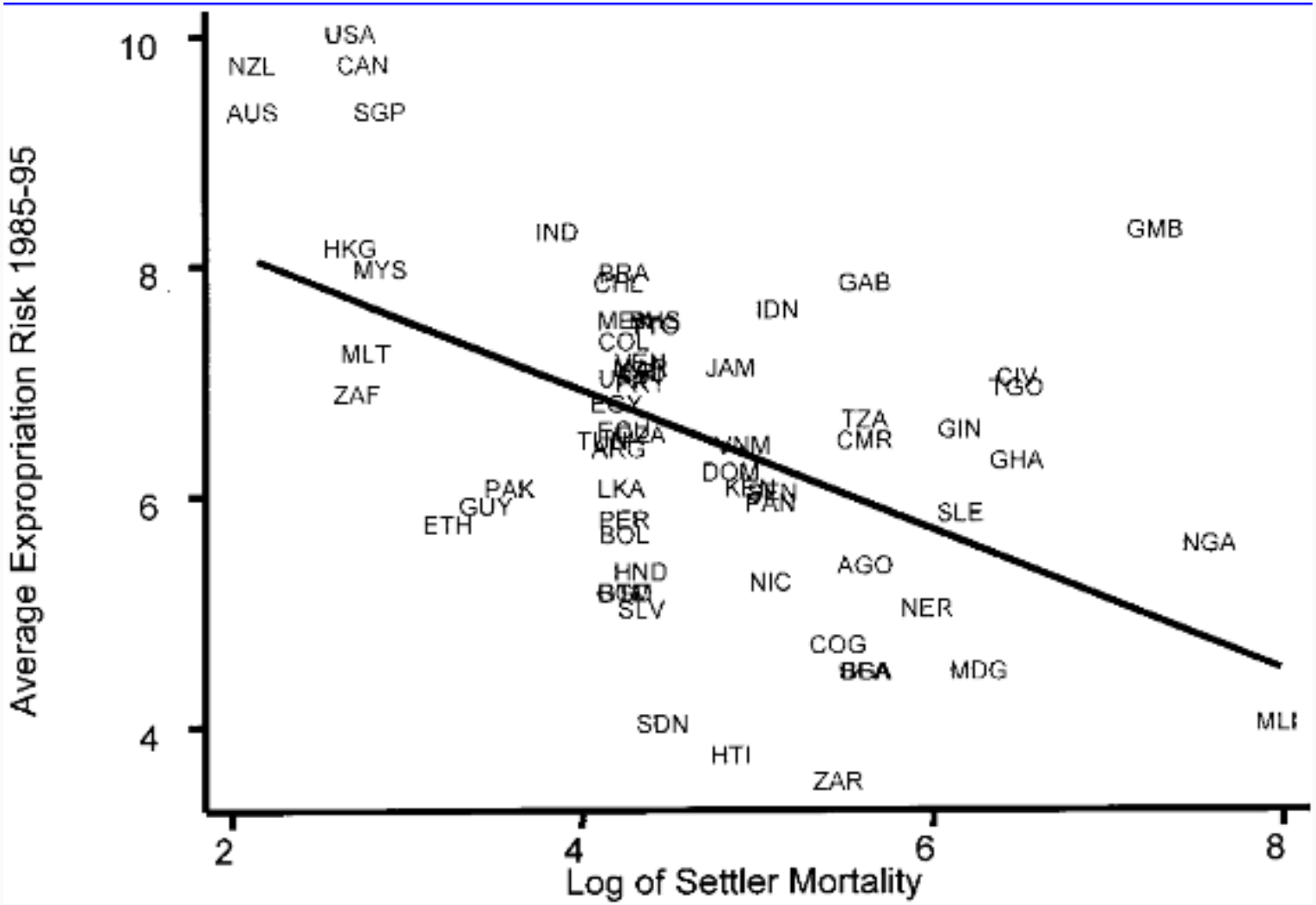
Settler Mortality: IV Relevance [SA8-Q1e]

- How would you write out the regression equation to estimate whether the first stage is strong?

$$C_i = \delta + \eta S_i + \mathbf{X_i}\theta + \nu_i$$

- What coefficient should we look at to evaluate the strength of the first stage? We look at $\hat{\eta}$: This should be very significant. The relevant F-statistic (with one instrument, this is the **square** of the t-statistic) should be at least 10.

Settler Mortality: IV Relevance [SA8-Q1e]



Settler Mortality: IV Reduced Form [SA8-Q1f]

- The authors document a strong ”reduced-form” relationship. Write out the regression equation used to estimate the reduced form

Settler Mortality: IV Reduced Form [SA8-Q1f]

- The authors document a strong "reduced-form" relationship. Write out the regression equation used to estimate the reduced form

$$\ln y_i = \mu + \lambda S_i + \mathbf{X_i}\kappa + e_i$$

Settler Mortality: IV Reduced Form [SA8-Q1f]

- The authors document a strong "reduced-form" relationship. Write out the regression equation used to estimate the reduced form

$$\ln y_i = \mu + \lambda S_i + \mathbf{X_i}\kappa + e_i$$

- Why don't we just estimate the reduced form?

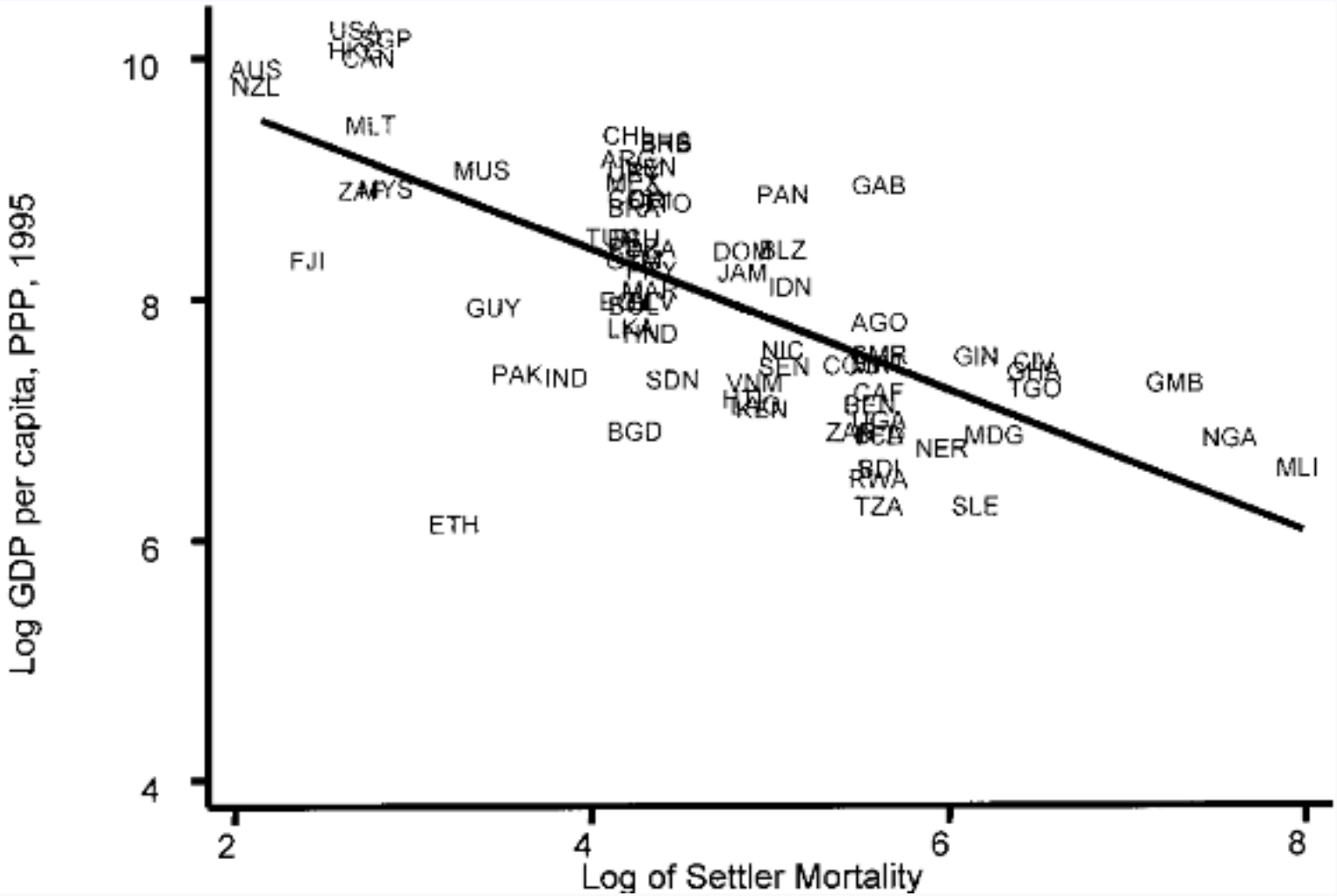
Settler Mortality: IV Reduced Form [SA8-Q1f]

- The authors document a strong "reduced-form" relationship. Write out the regression equation used to estimate the reduced form

$$\ln y_i = \mu + \lambda S_i + \mathbf{X_i}\kappa + e_i$$

- Why don't we just estimate the reduced form? We are not primarily interested in the effect of settler mortality, but we use settler mortality as a way to find out the causal effect of **institutions** on development outcomes.

Settler Mortality: IV Reduced Form [SA8-Q1f]



Settler Mortality: Results [SA8-Q1g]

	OLS			2SLS		
Average protection against expropriation risk (s_i)	0.52	0.47	0.41	0.94	1.00	1.10
	(0.06)	(0.06)	(0.06)	(0.16)	(0.22)	(0.46)
Latitude		0.89	0.92		-0.65	- 1.20
		(0.49)	(0.63)		(1.34)	(1.8)
Asia dummy			-0.60			-1.10
			(0.23)			(0.52)
Africa dummy			-0.90			-0.44
			(0.17)			(0.42)
"Other" continent dummy			-0.04			-0.99
			(0.32)			(1.00)

- Interpret the results of the OLS regressions
- Interpret the results of the 2SLS (IV) regressions
- How do the OLS and 2SLS estimates differ? What does this imply about the OLS regression from part b)?

Settler Mortality: Results [SA8-Q1g]

- OLS: An increase in the index of protection against expropriation by one unit is associated with around 40-50% higher GDP per capita (keeping latitude and continent fixed)

Settler Mortality: Results [SA8-Q1g]

- OLS: An increase in the index of protection against expropriation by one unit is associated with around 40-50% higher GDP per capita (keeping latitude and continent fixed)
- IV: An increase in the index of protection against expropriation by one unit leads to around 100% higher GDP per capita (keeping latitude and continent fixed)

Settler Mortality: Results [SA8-Q1g]

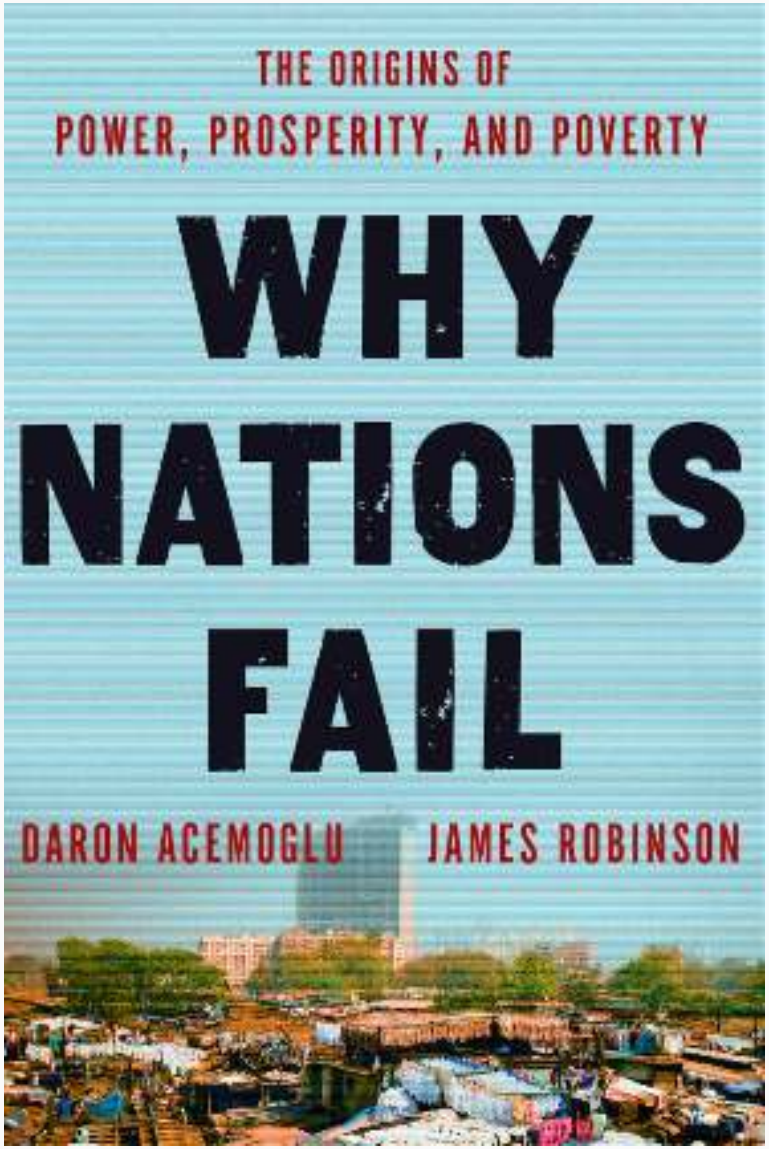
- OLS: An increase in the index of protection against expropriation by one unit is associated with around 40-50% higher GDP per capita (keeping latitude and continent fixed)
- IV: An increase in the index of protection against expropriation by one unit leads to around 100% higher GDP per capita (keeping latitude and continent fixed)
- The IV estimates are larger than the OLS estimates. Negative OVB or reverse causality should make the IV estimate smaller than the OLS estimate, so they may not be important here. The OLS estimate could be biased towards zero because of measurement error. In addition, the countries for which settler mortality had a big effect on today's institutions may be the countries for which institutions had the biggest effect on GDP (LATE).

Settler Mortality: Results [SA8-Q1g]

All the ways IV estimates can be different from OLS estimates

- **OVB:** If OLS had omitted variable bias, and our instrument is valid, then the IV estimate should be different. We can check whether this is plausible with the OVB formula
- **Measurement error:** If we have random measurement error in the independent variable (X), then we can use IV to overcome this. In that case, the IV coefficient will be larger (in absolute value) than the OLS coefficient
- **LATE:** OLS gives us ATT + Selection Bias, while IV gives us the treatment effect on the compliers (LATE). The ATT may be different from LATE, even without selection bias.
- **Invalid IV:** Hard to determine what exactly is going on
- **Sampling variation**: This can just happen by chance

Want to know more?



Any questions?

… Remember – Every question is useful!

LATE [SA8-Q2]

LATE: IV gives us the treatment effect for the compliers

Potential outcomes! (unobserved)		***Does not get voucher (Z=0)***	
Gets voucher (Z=1)		*Eats chocolate (D=1)*	*Does not eat chocolate (D=0)*
	Eats chocolate (D=1)	Always-takers: $E(D\|Z=1)=E(D\|Z=0)=1$ ➔ $E(Y\|Z=1)=E(Y\|Z=0)$	Compliers
	Does not eat chocolate (D=0)	Defiers	Never-takers: $E(D\|Z=1)=E(D\|Z=0)=0$ ➔ $E(Y\|Z=1)=E(Y\|Z=0)$

Health Insurance and LATE [SA8-Q2a]

Suppose individuals are eligible for a public health insurance program and are randomly assigned a priority number Z, which influences how likely they are to be enrolled in the program. Let β_{1i} represent the causal effect for individual i of enrolling in public health insurance on health outcomes, and let π_{1i} represent the effect of Z on the likelihood that individual i enrolls.

Half the individuals are aware that their health outcomes would improve from the program and thus may decide to enroll if given the option. For this population, would β_{1i} be positive, zero, or negative? What about π_{1i} ? Call these β_{1i}^{A} and π_{1i}^{A}.

Never takers? [SA8-Q2b]

The other half, perhaps those who are healthier, know their health outcomes would not improve from enrolling in health insurance. These individuals would not enroll even if they were admitted. Are β_{1i} and π_{1i} positive, zero, or negative for this group? Call these β_1^B and π_1^B.

IV and ATE [SA8-Q2d]

If we use Z as an instrument for enrollment, would the resulting estimate identify the average treatment effect (ATE) of health insurance on health outcomes? Please explain your reasoning.

IV and ATE [SA8-Q2d]

Solution No. In general, the 2SLS estimate identifies a local average treatment effect (LATE). The LATE equals the ATE only if at least one of the following cases holds: (i) the treatment effect is the same for all individuals, (ii) the instrument affects all individuals equally, or (iii) heterogeneity in the treatment effect and heterogeneity in the effect of the instrument are uncorrelated.

Based on (a) and (b), the treatment effect and the effect of the instrument are different for the two groups, ruling out cases (i) and (ii). For group (a), $\beta_{1i}^{A} > 0$ and $\pi_{1i}^{A} > 0$, whereas for group (b) $\beta_{1i}^{B} \leq 0$ and $\pi_{1i}^{B} = 0$. This implies that $\mathbf{cov}\left(\beta_{1i}, \pi_{1i}\right) > 0$, ruling out case (iii).

Calculating LATE [SA8-Q2e]

Derive and interpret the local average treatment effect (LATE).
How does the LATE compare to the ATE?

Calculating LATE [SA8-Q2e]

Solution

When both the treatment effect and the effect of the instrument are heterogeneous, $\widehat{\beta}^{IV}$ estimates the following:

$$\widehat{\beta}^{IV} = \frac{E\left[\beta_{1i}\pi_{1i}\right]}{E\left[\pi_{1i}\right]} = \text{ LATE}$$

Let's evaluate this:

$$E\left[\pi_{1i}\right] = 0.5 * \pi_1^A + 0.5 * \pi_1^B = 0.5 * \pi_1^A$$

$$E\left[\beta_{1i}\pi_{1i}\right] = 0.5 * \beta_1^A\pi_1^A + 0.5 * \beta_1^B\pi_1^B = 0.5 * \beta_1^A\pi_1^A$$

Plugging these into the expression for the LATE, we get that the LATE identifies β_1^A, the causal effect for individuals who might enroll in health insurance. It gives zero weight to individuals who would never enroll. Because in this case $\beta_1^A > \beta_1^B$, the LATE exceeds the average treatment effect.

Any questions?

… Remember – Every question is useful!

Differences in Differences

Econ 140, Section 9

Jonathan Old

Roadmap

1. Motivation

2. SA10-Q2a: Calculating Diff-in-Diff

3. Rainfall IV: Checking relevance with DiD [SA10-Q3]

Any questions?

… Remember – Every question is useful!

Motivation

Motivation

- In the first part of the course, we talked about comparison between groups or other units (cross-sectional data)
- But we have also seen some comparisons over time: Gas price example in SA8 (time-series data)
- Both of these have very big and obvious problems, but we can use them **together** using (panel data) in a powerful tool: Differences-in-differences!

How to get the causal effect of a treatment

In 2021, UC Berkeley offered free mental health coachings to students with pre-existing mental health issues. We want to evaluate the effect of this policy and we collect a depression score (0-10) for all students at UC Berkeley.

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can do several comparisons:

How to get the causal effect of a treatment II

	2020	2022
Free Mental Health: Treated		6
No Free Mental Health: Untreated		5

We can do several comparisons:

- Comparison 1: Compare treated and non-treated group after the intervention

How to get the causal effect of a treatment II

	2020	2022
Free Mental Health: Treated		6
No Free Mental Health: Untreated		5

We can do several comparisons:

- Comparison 1: Compare treated and non-treated group after the intervention
- Estimated treatment effect? $6 - 5 = 1$

How to get the causal effect of a treatment II

	2020	2022
Free Mental Health: Treated		6
No Free Mental Health: Untreated		5

We can do several comparisons:

- Comparison 1: Compare treated and non-treated group after the intervention
- Estimated treatment effect? $6 - 5 = 1$
- Problems?

How to get the causal effect of a treatment II

	2020	2022
Free Mental Health: Treated		6
No Free Mental Health: Untreated		5

We can do several comparisons:

- Comparison 1: Compare treated and non-treated group after the intervention
- Estimated treatment effect? $6 - 5 = 1$
- Problems?
- **Identifying assumption**: The average difference between groups is due to the treatment only. Without the treatment, the average outcome of the treated group would have been equal to the average outcome of the control group.

How to get the causal effect of a treatment III

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated		

We can do several comparisons:

- Comparison 2: Compare treated group before and after the intervention

How to get the causal effect of a treatment III

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated		

We can do several comparisons:

- Comparison 2: Compare treated group before and after the intervention
- Estimated treatment effect? $6 - 6 = 0$

How to get the causal effect of a treatment III

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated		

We can do several comparisons:

- Comparison 2: Compare treated group before and after the intervention
- Estimated treatment effect? $6 - 6 = 0$
- Problems?

How to get the causal effect of a treatment III

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated		

We can do several comparisons:

- Comparison 2: Compare treated group before and after the intervention
- Estimated treatment effect? $6 - 6 = 0$
- Problems?
- Identifying assumption: The average difference across time is due to the treatment only. Without the treatment, the average outcome of the treated group would not have changed.

How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can do several comparisons:

- Comparison 3: Compare treated to untreated group, before and after the intervention. **Differences in differences!**

How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can do several comparisons:

- Comparison 3: Compare treated to untreated group, before and after the intervention. **Differences in differences!**
- Estimated treatment effect?
  $(6 - 6) - (5 - 4) = (6 - 5) - (6 - 4) = -1$

How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can do several comparisons:

- Comparison 3: Compare treated to untreated group, before and after the intervention. **Differences in differences!**
- Estimated treatment effect?
  $(6 - 6) - (5 - 4) = (6 - 5) - (6 - 4) = -1$
- Identifying assumption: Parallel trends: Without the treatment, the average increase in the outcome of the treated would have been the same as the average increase in the outcome of the untreated.

Sidenote

... We can NOT observe the identifying assumption! We can find evidence for or against it, but we can never be sure!

How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can calculate the DiD estimate in two ways (here: superscripts stand for treatment/control group, subscripts denote unit i and time $t \in \{0, 1\}$)

$$DiD = E[\ \underbrace{\left(Y_{i1}^{T} - Y_{i0}^{T}\right)}_{\text{Change for treated}} - \underbrace{\left(Y_{i1}^{C} - Y_{i0}^{C}\right)}_{\text{Change for untreated}}\ ]$$

$$= E[\ \underbrace{\left(Y_{i1}^{T} - Y_{i1}^{C}\right)}_{\text{After-difference}} - \underbrace{\left(Y_{i0}^{T} - Y_{i0}^{C}\right)}_{\text{Before-difference}}\ ]$$

DiD and Potential Outcomes [SA10-Q1]

Let us look at the second one in more detail:

$$
\begin{aligned}
DiD &= E[\left(Y_{i1}^T - Y_{i1}^C\right) - \left(Y_{i0}^T - Y_{i0}^C\right)] \\
&= \underbrace{\left(E[Y_{i1}^T] - E[Y_{i1}^C]\right)}_{\text{After-difference}} - \underbrace{\left(E[Y_{i0}^T] - E[Y_{i0}^C]\right)}_{\text{Before-difference}}
\end{aligned}
$$

Compare this with out formula for selection bias (in time period 1):

$$
\begin{aligned}
\Delta &= E[Y_{i1}^T] - E[Y_{i1}^C] \\
&= \underbrace{E[\mathrm{Y}_{i1}^T(1) - \mathrm{Y}_{i1}^T(0)]}_{\text{ATT}} + \underbrace{\left(E[\mathrm{Y}_{i1}^T(0)] - E[\mathrm{Y}_{i1}^C(0)]\right)}_{\text{Selection Bias}}
\end{aligned}
$$

DiD and Potential Outcomes [SA10-Q1]

Let us look at the second one in more detail:

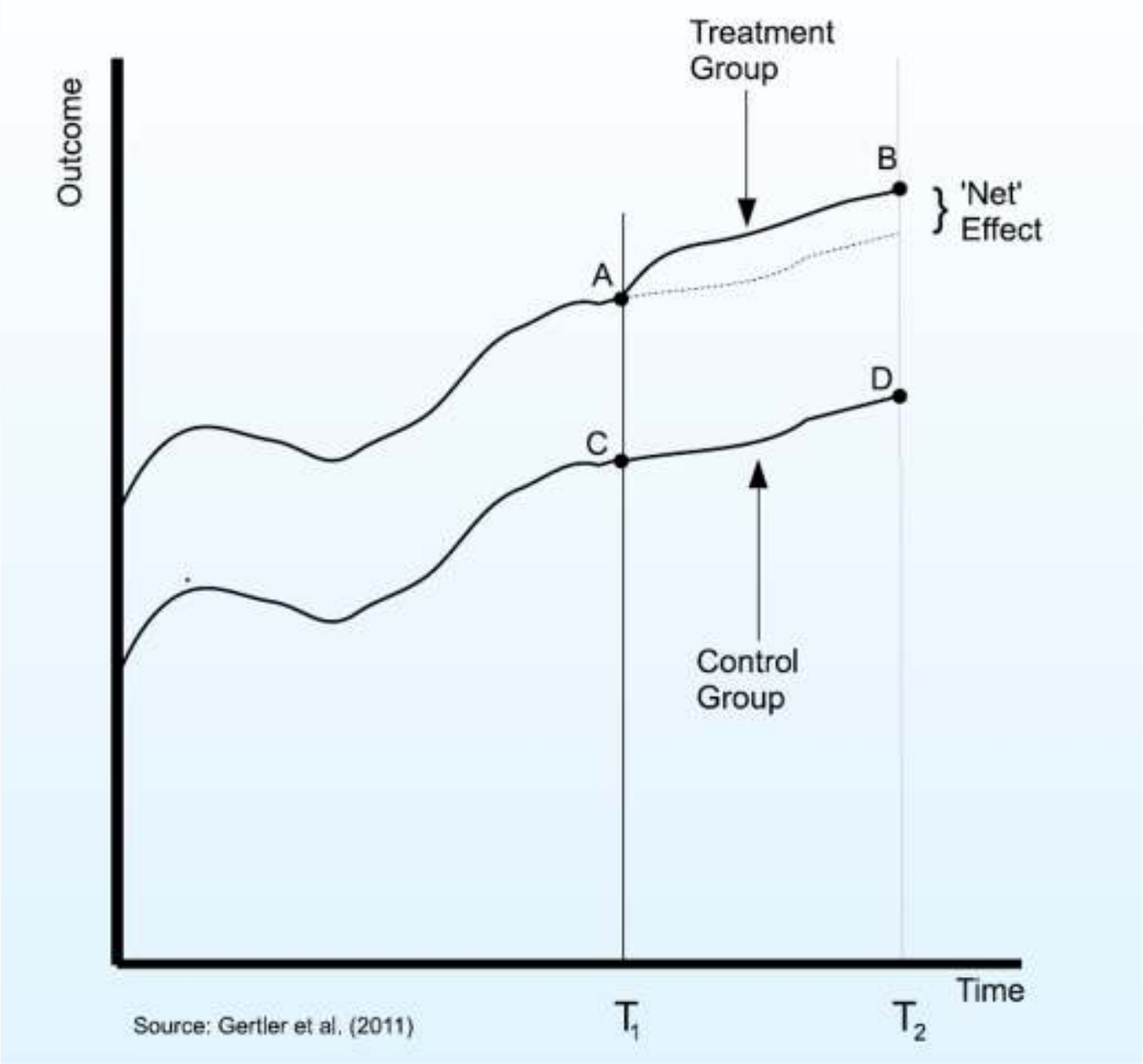
$$DiD = E[\left(Y_{i1}^T - Y_{i1}^C\right) - \left(Y_{i0}^T - Y_{i0}^C\right)]$$
$$= \underbrace{\left(E[Y_{i1}^T] - E[Y_{i1}^C]\right)}_{\text{After-difference}} - \underbrace{\left(E[Y_{i0}^T] - E[Y_{i0}^C]\right)}_{\text{Before-difference}}$$

Compare this with out formula for selection bias (at $t = 1$):

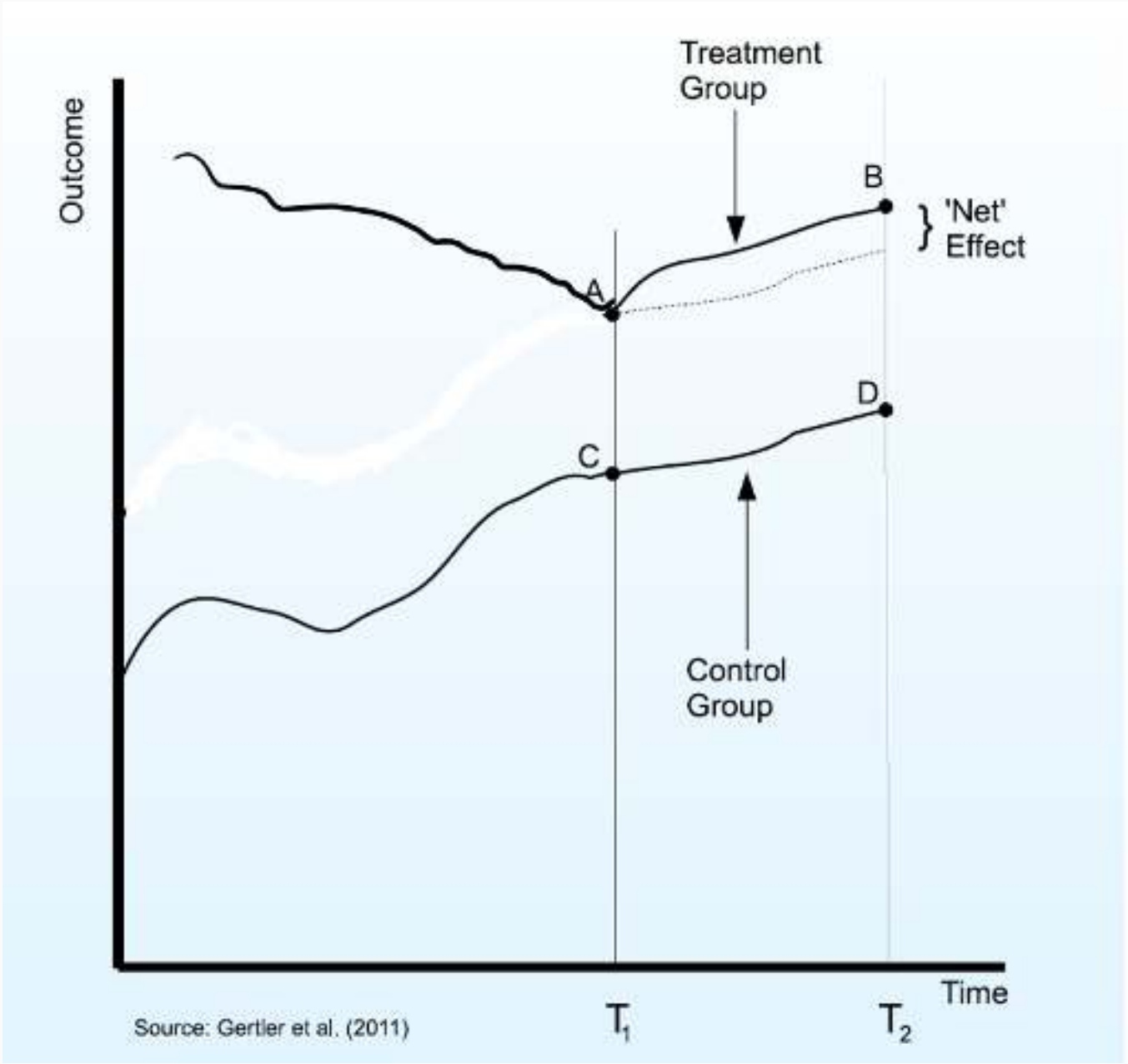
$$\Delta = E[Y_{i1}^T] - E[Y_{i1}^C]$$
$$= \underbrace{E[Y_{i1}^T(1) - Y_{i1}^T(0)]}_{\text{ATT}} + \underbrace{\left(E[Y_{i1}^T(0)] - E[Y_{i1}^C(0)]\right)}_{\text{Selection Bias}}$$

So DiD=ATT if the before-difference is equal to the selection bias in period 1: If the treatment group hadn't gotten the treatment (potential outcome = 0), the difference to the control group would have been the same as in period 0.

Parallel trends assumption



(Potential) violation of a parallel trends assumption



Estimating DiD with regressions

We can set up a simple linear regression to estimate a DiD model:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

	2020	2022
Free Mental Health: Treated		
No Free Mental Health: Untreated	α	

Estimating DiD with regressions

We can set up a simple linear regression to estimate a DiD model:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

	2020	2022
Free Mental Health: Treated		
No Free Mental Health: Untreated	α	$\alpha + \gamma$

Estimating DiD with regressions

We can set up a simple linear regression to estimate a DiD model:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

	2020	2022
Free Mental Health: Treated	$\alpha + \beta$	
No Free Mental Health: Untreated	α	$\alpha + \gamma$

Estimating DiD with regressions

We can set up a simple linear regression to estimate a DiD model:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

	2020	2022
Free Mental Health: Treated	$\alpha + \beta$	$\alpha + \beta + \gamma + \delta$
No Free Mental Health: Untreated	α	$\alpha + \gamma$

Estimating DiD with regressions

We can set up a simple linear regression model to get the DiD estimate:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

With our data, we would get:

- $\alpha = 4$
- $\beta = 2$
- $\gamma = 1$
- $\delta = -1$

Any questions?

… Remember – Every question is useful!

SA10-Q2a: Calculating Diff-in-Diff

SA10-Q2a

	Area: **Sure Start**	**Non-Sure Start**	$\sum$
Year: 2006	40	45	-5
2008	45	47	-2
$\sum$	5	2	**3**

SA10-Q2a

	Area: **Sure Start**	**Non-Sure Start**	$\sum$
Year: 2006	40	45	-5
2008	45	47	-2
$\sum$	5	2	**3**

The DiD estimate is:

SA10-Q2a

	Area: **Sure Start**	**Non-Sure Start**	$\sum$
Year: 2006	40	45	-5
2008	45	47	-2
$\sum$	5	2	**3**

The DiD estimate is:

$$(45 - 40) - (47 - 45) = (45 - 47) - (40 - 45) = 3$$

The ”percentage” DiD estimate is:

SA10-Q2a

	Area: **Sure Start**	**Non-Sure Start**	$\sum$
Year: 2006	40	45	-5
2008	45	47	-2
$\sum$	5	2	**3**

The DiD estimate is:

$$(45 - 40) - (47 - 45) = (45 - 47) - (40 - 45) = 3$$

The ”percentage” DiD estimate is:

$(45{-}40)/40{-}(47{-}45)/45 = (12.5\%{-}4.4\%) = 8.1$

SA10-Q2a

	Area: **Sure Start**	**Non-Sure Start**	$\sum$
Year: 2006	40	45	-5
2008	45	47	-2
$\sum$	5	2	**3**

The DiD estimate is:

$$(45 - 40) - (47 - 45) = (45 - 47) - (40 - 45) = 3$$

The ”percentage” DiD estimate is:

$(45-40)/40-(47-45)/45 = (12.5\%-4.4\%) = 8.1$ percentage points

SA10-Q2b: Checking parallel trends

	Area: **Sure Start**	**Non-Sure Start**	$\sum$
Year: 2006	40	45	-5
2008	45	47	-2
$\sum$	5	2	**3**

Can you think of a possible violation of parallel trends?

SA10-Q2b: Checking parallel trends

	Area: **Sure Start**	**Non-Sure Start**	$\sum$
Year: 2006	40	45	-5
2008	45	47	-2
$\sum$	5	2	**3**

Can you think of a possible violation of parallel trends?

- Parental education: If parents in the Sure Start program have lower education, their children **would have** had a **slower increase** in the absense of the treatment.
- Other interventions: Maybe children in the Sure Start program also got other treatments, e.g. subsidized kindergarten meals.

DiD with controsl (SA10-Q2c)

We can also include control variables in a DiD regression:

$$Y_{it} = \beta_0 + \beta_1 T_i + \beta_2 \text{ Post }_t + \beta_3 (T_i * \text{ Post }_t) + \beta_4 \text{ Depression }_{it} + u_{it}$$

The key is:

- Anything that is constant across time and that does not affect the difference between T and C is already accounted for by DiD
- For anything that **changes over time** or **affects the difference between T and C**, we need to include a control variable

Rainfall IV: Checking relevance with DiD [SA10-Q3]

Rainfall and Income

We want to know the effect of income on education and estimate

$$Educ_i = \beta_0 + \beta_1 \text{ Income }_i + \beta_2 \text{ Pop }_i + \beta_3 \text{ School }_i + \beta_4 \text{ Age }_i + u_i$$

Explain what econometric problem is likely to arise that leads to biased and inconsistent estimates as a result of including Income as a regressor in the education regression as is done above.

IV regression

We want to know the effect of income on education and estimate

$$Educ_i = \beta_0 + \beta_1 \text{ Income }_i + \beta_2 \text{ Pop }_i + \beta_3 \text{ School }_i + \beta_4 \text{ Age }_i + u_i$$

You learn from Ghana's Minister of Agriculture that the country's citizens derive the bulk of their income from agriculture. As a result, you cleverly infer that average annual rainfall (Rainfall) may be a good instrument for income. You recall from your econometrics course that an instrument can be used in a procedure called Two Stage Least Squares that is designed to solve this econometric problem. Describe carefully the first of the two stages.

Checking Relevance with DiD

You want to check the Minister's suggestion that rainfall has an impact on incomes in Ghana. You have information on average annual incomes in 1996 and 1997 for two regions: the "coastal region," which had the same precipitation level in both years, and the "hill region," which experienced a 30% increase in rainfall. Comparing 1996 and 1997, income in the coastal region fell from 124 to 104, while income in the hill region fell from 98 to 96 . You also recall from your econometrics course that this situation might represent a "natural" or "quasi" experiment, allowing you to estimate the "treatment effect" of rainfall. Perform a difference in differences analysis of the effect of rainfall on average income. Summarize the analysis in a table.

DiD summary

Region	Rainfall	1996	1997
Coast (control)	No change	124	104
Hill (treatment)	+30%	98	96

This is equivalent to estimating:

$$I_{it} = \beta_0 + \beta_1 G_i + \beta_2 D_t + \beta_3 G_i \times D_t + u_i$$

where $G_i = 1$ if village is in Hills and $G_i = 0$ if village is on the Coast; $D_t = 1$ if year is 1997 and $D_t = 0$ if year is 1996. The differences in differences estimate of the rainfall effect is the OLS of β_3.

Exercises and Questions around IV

Econ 140, Section 10

Jonathan Old

Roadmap

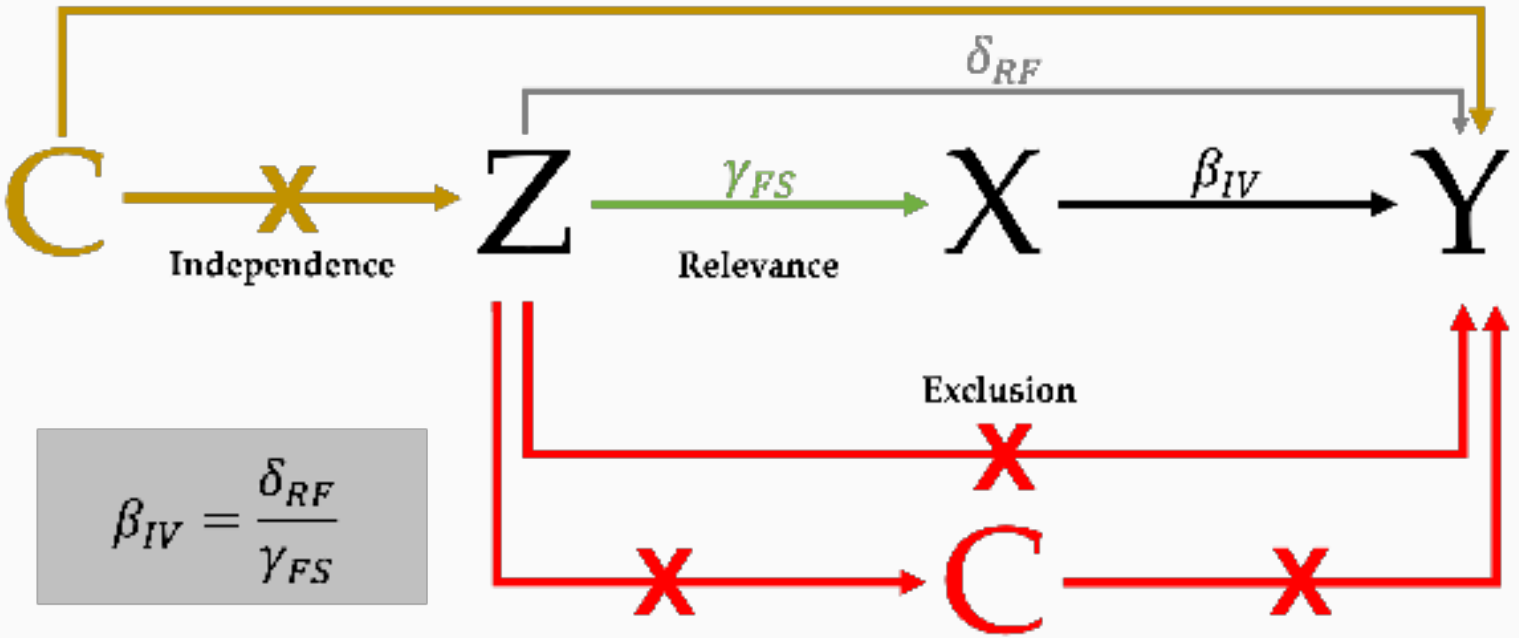
1. Recap: IV
2. Group work
3. IV Exam Exercise
4. Questions

Recap: IV

Recap: IV summary

We need the following three assumptions for IV to work:

1. **Relevance**: *Z* must truly affect *X*
2. **Independence/Exogeneity**: *Z* is as good as randomly assigned
3. **Exclusion Restriction**: The **only** way that *Z* affects *Y* is via *X*.



$$\beta_{IV} = \frac{\delta_{RF}}{\gamma_{FS}}$$

How IV estimates can be different from OLS estimates

- **OVB:** If OLS had omitted variable bias, and our instrument is valid, then the IV estimate should be different. We can check whether this is plausible with the OVB formula
- **Measurement error:** If we have random measurement error in the independent variable (X), then we can use IV to overcome this. In that case, the IV coefficient will be larger (in absolute value) than the OLS coefficient
- **LATE:** OLS gives us ATT + Selection Bias, while IV gives us the treatment effect on the compliers (LATE). The ATT may be different from LATE, even without selection bias.
- **Invalid IV:** Hard to determine what exactly is going on
- **Sampling variation**: This can just happen by chance

IV and controls

- There is a catch in IV: independence (exogeneity) and exclusion restriction only need to hold **conditionally**
- This means: If we **know** C, we can control for it in the IV regression, and the concerns go away
- But then, we have the old OLS problems again
- So we should only do this in cases where we really know all the C's.

Group work

Group work

Group 1: We are interested in the effect of being in the army on crime. We instrument being in the army with a lottery (paper)

Group 2: We are interested in the effect of income on conflict. We instrument income with rainfall (paper)

Group 3: We are interested in the effect of air pollution on mortality. We instrument local air pollution with wind direction (paper)

1. Relevance: Z must truly affect X
2. Independence/Exogeneity: Z is as good as randomly assigned
3. Exclusion restriction: The **only** way that Z affects Y is via X

Your job: Discuss whether these assumptions hold!

IV Exam Exercise

IV Exam exercise i

The land of Econometrea consists of 34 islands. Each island has their own university and students attend uni on their home island. Students in Econometrea either have to attend lectures in person or watch them online. Priyanka has obtained data for the students from all the universities and would like to study the effect of watching lectures online on the students' exam scores. For each student she has the exam score (0 - 100, 70+ is a first, 40 and below a fail), which she uses as her left hand side variable, the fraction of lectures watched online, and how on how many days the student visited the library each week. Priyanka obtains the following regression results:

IV Exam exercise ii

	Estimation Method				
	OLS	IV	IV	IV	IV
Regressor	(1)	(2)	(3)	(4)	(5)
Fraction online	−4.91	−2.08	−0.56	−0.75	6.09
	(0.11)	(0.22)	(0.40)	(0.90)	(3.22)
Library visits			0.91		2.09
			(0.08)		(0.43)
Sample	All	All	All	Unis with lottery	Unis with lottery

All regressions also contain a constant term and a dummy variable for each but one of the universities. Robust standard errors are reported in parentheses.

IV Exam exercise iii

(a) What is the interpretation of the coefficient in column (1). If this were a causal effect, would it be big or small? Explain whether this estimate is likely to have a causal interpretation.
(b) Wei Min observes that some halls of residence are close to the lecture rooms while others are further away and suggests to use the distance of halls from lecture halls as an instrumental variable for the fraction of lectures watched online. Results are in column (2). Explain which assumptions need to be satisfied for the IV to deliver better causal estimates than OLS here. Discuss the validity of the

IV Exam exercise iv

assumptions in this case.

	Estimation Method				
	OLS	IV	IV	IV	IV
Regressor	(1)	(2)	(3)	(4)	(5)
Fraction online	−1.91	−2.08	−0.56	−0.75	6.09
	(0.11)	(0.22)	(0.40)	(0.90)	(3.22)
Library visits			0.91		2.09
			(0.08)		(0.43)
Sample	All	All	All	Unis with lottery	Unis with lottery

All regressions also contain a constant term and a dummy variable for each but one of the universities. Robust standard errors are reported in parentheses.

IV Exam exercise v

Solution

- Students who watch all lectures online, on average, have around 4.9 fewer points on the exam than students who went to all lecturs in person. This is a moderately strong effect, but likely not causal (think of omitted variables, such as student motivation)

IV Exam exercise vi

Solution

- **Relevance:** Proximity to lecture halls must correlate strongly with lecture attendance. This is plausible from our experience, and we could test this using an F-test.
- **Independence/Exogeneity:** Proximity must be as good as randomly allocated. This might not hold here, for example, if more motivated students live closer to campus.
- **Exclusion restruction:** Proximity to lecture halls only influences exam scores because it has an effect on lecture attendance. This can be violated if proximity also affects the probability to go and study in the library, or if residence halls closer to campus are also closer to bars

IV Exam exercise vii

(c) Alma points out to Wei Min that students can pick which hall they want to live in and that some halls are located in Study Village, close to lecture halls and the library while others are located further away in Party Town, surrounded by pubs. How does this information affect your assessment of the IV strategy?
(d) Manuel realises that the data also include a variable for the number of times a student has checked into the library per week. He suggests to rerun the instrumental variables regression adding this variable as a control. Results for this regression are displayed in column (3). Assess Manuel's strategy.

IV Exam exercise viii

Solution

- This would be a violation of independence/exogeneity: The students who live closer to the campus are just inherently different than students who live far away
- Library visits are a bad control: They are another outcome of the instrument, and so controlling for them introduces additional bias.

IV Exam exercise ix

	Estimation Method				
	OLS	IV	IV	IV	IV
Regressor	(1)	(2)	(3)	(4)	(5)
Fraction online	−4.91	−2.08	−0.56	−0.75	6.09
	(0.11)	(0.22)	(0.40)	(0.90)	(3.22)
Library visits			0.91		2.09
			(0.08)		(0.43)
Sample	All	All	All	Unis with lottery	Unis with lottery

All regressions also contain a constant term and a dummy variable for each but one of the universities. Robust standard errors are reported in parentheses.

IV Exam exercise x

(e) Jenny notices that there are five universities which assign students to their halls of residence by a lottery. She suggests to run the IV model from columns (2) and (3) for the subsample of students from these universities only. Results are displayed in columns (4) and (5). Assess Jenny's regressions.
(f) Drawing on the results in the table above, what have you learned from this exercise about the causal effect of watching lectures online on students' exam results?

IV Exam exercise xi

Solution

- For these universities, the distance is as good as randomly assigned, and so we have no violations of the independence/exogeneity restriction. However, the exclusion restriction may still not hold.
- In a simply OLS regression, it may look as if watching lectures online causes worse grades. However, we can correct for this using a valid IV, and find that the effect is only very small and statistically insignificant.

Questions

From Diff in Diff to Panel Data

Econ 140, Section 11

Jonathan Old

Roadmap

1. Recap: DiD

2. Making DiD more general: Panel Data estimation

3. Section Assignment

Recap: DiD

How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can do several comparisons:

- Comparison 3: Compare treated to untreated group, before and after the intervention. **Differences in differences!**

How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can do several comparisons:

- Comparison 3: Compare treated to untreated group, before and after the intervention. **Differences in differences!**
- Estimated treatment effect?
  $(6 - 6) - (5 - 4) = (6 - 5) - (6 - 4) = -1$

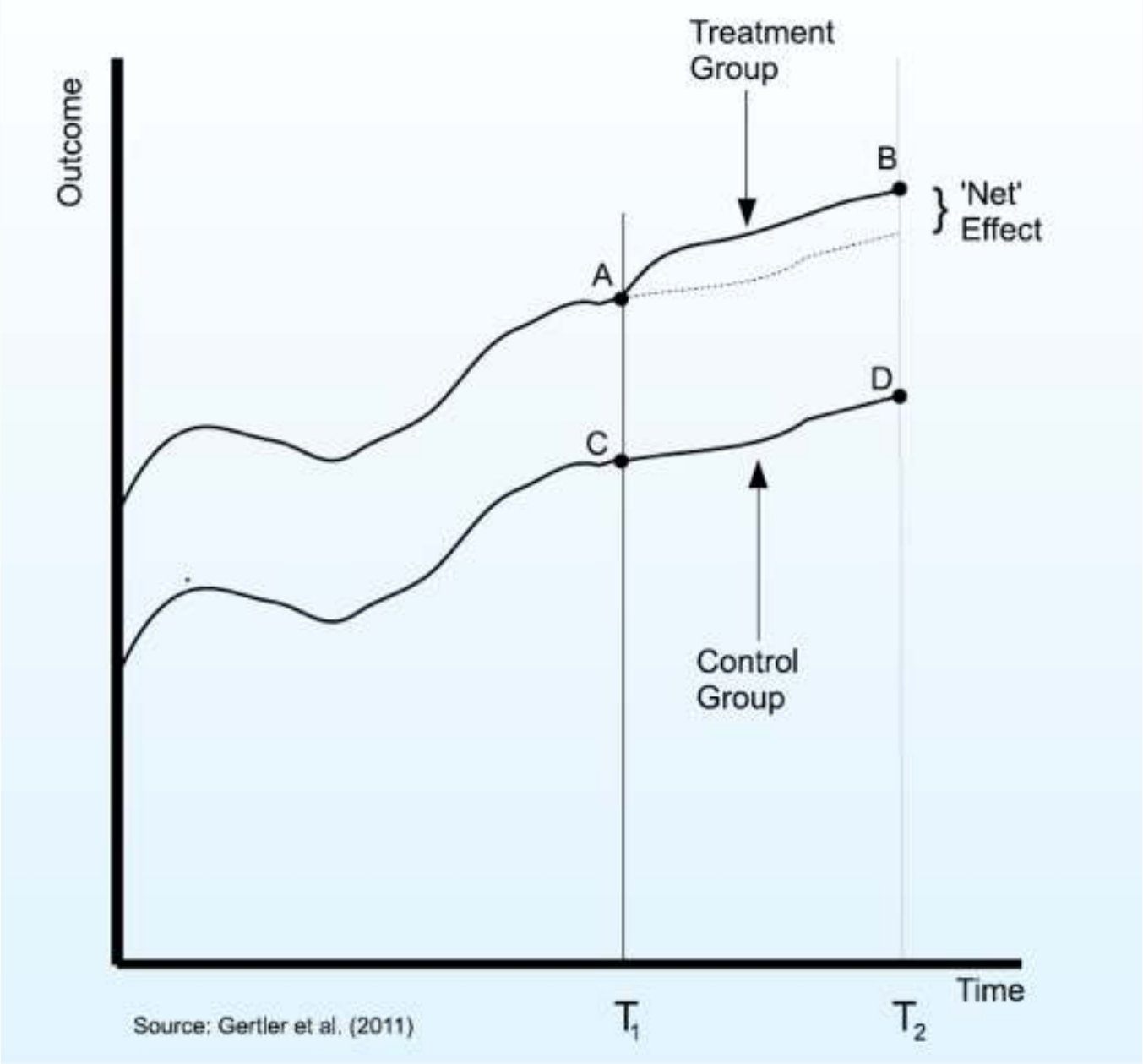
How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

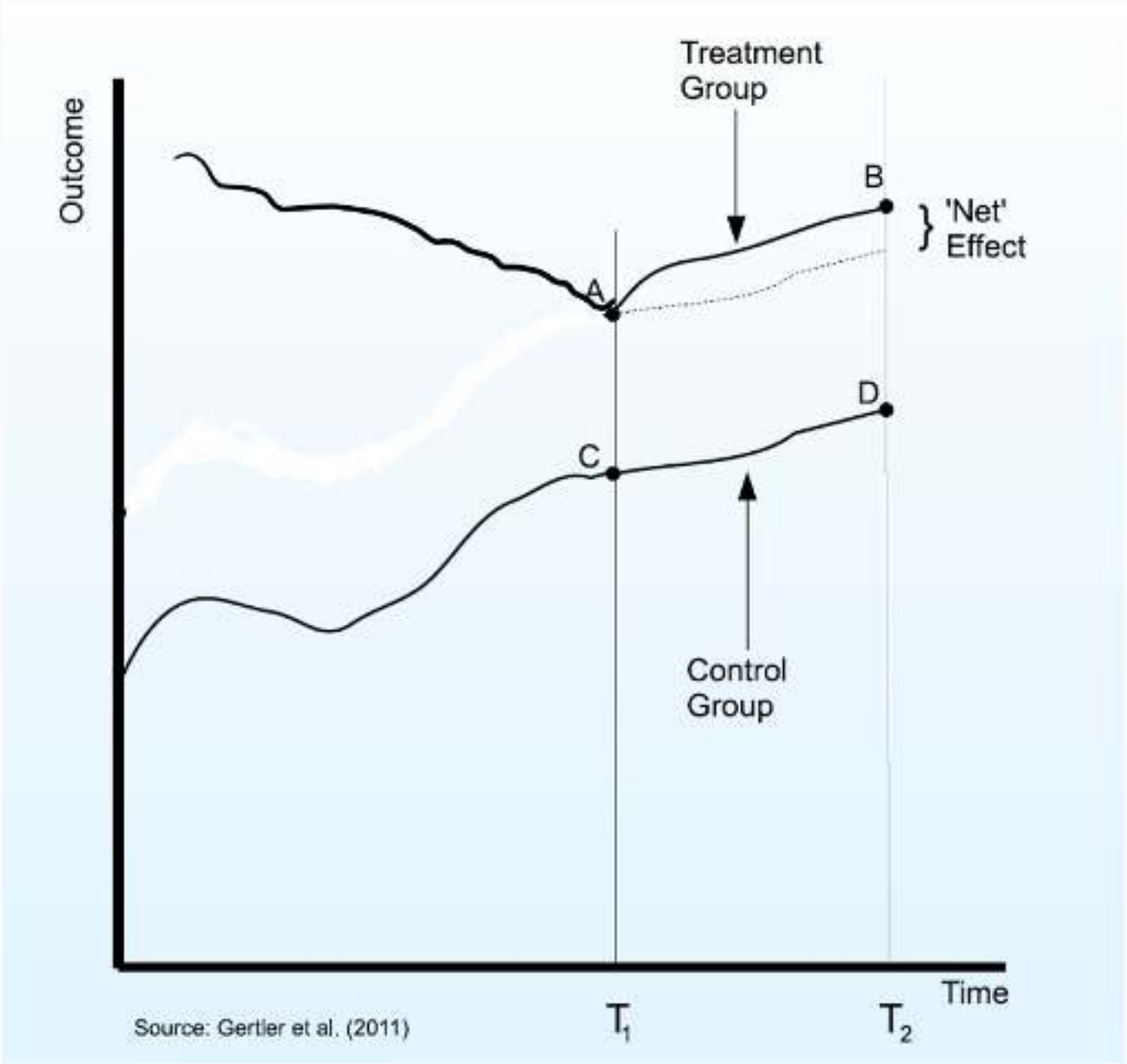
We can do several comparisons:

- Comparison 3: Compare treated to untreated group, before and after the intervention. **Differences in differences!**
- Estimated treatment effect?
  $(6 - 6) - (5 - 4) = (6 - 5) - (6 - 4) = -1$
- Identifying assumption: Parallel trends: Without the treatment, the average increase in the outcome of the treated would have been the same as the average increase in the outcome of the untreated.

Parallel trends assumption



(Potential) violation of a parallel trends assumption



Estimating DiD with regressions

We can set up a simple linear regression to estimate a DiD model:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

	2020	2022
Free Mental Health: Treated	$\alpha + \beta$	$\alpha + \beta + \gamma + \delta$
No Free Mental Health: Untreated	α	$\alpha + \gamma$

Any questions?

… Remember – Every question is useful!

Making DiD more general: Panel Data estimation

Limitations of the 2x2 DiD framework

In the simple 2x2 framework, we estimated:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

What are some potential limitations in this approach?

Intro to Panel Data i

A panel dataset may look like this:

	country	iso2c	iso3c	year	oda_per_cap	perc_below_pov_line	gdp_per_cap	inflation
316	Chile	CL	CHL	2015	3.0402700	0.4	13569.9478	4.34877353
317	Chile	CL	CHL	2016	9.8673527	NA	13785.6883	3.78619356
318	Chile	CL	CHL	2017	3.8277424	0.3	15045.5277	2.18271847
319	Chile	CL	CHL	2018	NA	NA	15795.7085	2.43488981
320	Chile	CL	CHL	2019	NA	NA	14631.9469	2.55754476
321	Chile	CL	CHL	2020	NA	0.7	13094.4595	3.04549085
322	Chile	CL	CHL	2021	NA	NA	16265.0960	4.52456838
323	China	CN	CHN	2015	-0.2219718	1.2	8016.4314	1.43702361
324	China	CN	CHN	2016	-0.5702808	0.8	8094.3634	2.00000162
325	China	CN	CHN	2017	-0.7090169	0.7	8816.9869	1.59313600
326	China	CN	CHN	2018	-0.5027741	0.4	9905.3420	2.07479040
327	China	CN	CHN	2019	0.4323084	0.1	10143.8382	2.89923416
328	China	CN	CHN	2020	0.4059415	NA	10408.6698	2.41942189
329	China	CN	CHN	2021	-0.4002945	NA	12556.3331	0.98101514
330	Colombia	CO	COL	2015	28.7760584	4.9	6228.4263	4.98983116
331	Colombia	CO	COL	2016	23.2331299	4.9	5938.4639	7.51346025
332	Colombia	CO	COL	2017	17.5871894	4.3	6450.3196	4.31431326
333	Colombia	CO	COL	2018	36.1416353	4.5	6782.0379	3.24056933
334	Colombia	CO	COL	2019	17.4085507	5.3	6458.0602	3.52301933

Intro to Panel Data ii

It may also look like this:

	Crime_2006	Crime_2007	Crime_2008	Crime_2009	Crime_2010	Crime_2011	Crime_2012	Crime_2013	Crime_2014	Crime_2015	Crime_2016	Crime_2017
New York	0,74075276	0,80312483	0,57854938	0,58480987	0,20097276	0,80826596	0,97070048	0,02437202	0,07831943	0,84572914	0,75749919	0,98948363
Los Angeles	0,81840795	0,63440595	0,75200179	0,45218138	0,54532883	0,0015679	0,39512876	0,41847691	0,66467241	0,01946605	0,55885996	0,27521867
Chicago	0,81682982	0,00089416	0,74575534	0,5521186	0,60461758	0,34694645	0,47874246	0,15925374	0,34667588	0,83944993	0,44825934	0,58746291
Houston	0,05885616	0,79349661	0,18033028	0,51317119	0,79642654	0,50269029	0,56157808	0,77173283	0,60797558	0,40461519	0,785332	0,6477959
Phoenix	0,2568483	0,28097691	0,82722175	0,11994755	0,884851	0,83800072	0,08978678	0,21613254	0,97306225	0,57291156	0,27341507	0,20641572
Philadelphia	0,86364857	0,30478317	0,07818559	0,82930161	0,28324875	0,60204952	0,4467311	0,1165021	0,15597767	0,36430802	0,38034388	0,37091781
San Antonio	0,98256294	0,13215208	0,55220195	0,25274349	0,22117177	0,63049228	0,74109619	0,17937274	0,50914785	0,68773119	0,17854372	0,50634586
San Diego	0,26052609	0,90551179	0,96145068	0,04932108	0,55859179	0,57753748	0,38254621	0,36728436	0,39429268	0,40312699	0,55506545	0,51665312
Dallas	0,85045327	0,33482019	0,15361924	0,74231968	0,55762722	0,05057305	0,57960707	0,71774087	0,63565644	0,04305477	0,06451316	0,27719973
San Jose	0,75452737	0,21462214	0,24820183	0,2727016	0,00127307	0,36822105	0,32304996	0,9116018	0,86006177	0,05904204	0,1124391	0,83247218
Austin	0,45398495	0,27082977	0,89920786	0,82480029	0,52877813	0,9452614	0,82578514	0,828743	0,20606871	0,4439492	0,38857348	0,8154967
Jacksonville	0,18638302	0,66744344	0,38392491	0,83800597	0,30374371	0,16785163	0,53614039	0,15558203	0,14260844	0,51748284	0,07054862	0,70962828
Fort Worth	0,13254183	0,78355507	0,83321159	0,33460775	0,00772913	0,78127289	0,61026409	0,25552213	0,27373821	0,78588972	0,80784366	0,1307398
Columbus	0,21457262	0,59542339	0,65347292	0,16532587	0,35580394	0,7113844	0,72036005	0,39323372	0,73022336	0,32649945	0,95442814	0,44923343
Indianapolis	0,97993863	0,2017115	0,58883837	0,54914541	0,22495244	0,58047281	0,01984661	0,88891756	0,34920826	0,326581	0,11407985	0,76592867
Charlotte	0,13043704	0,44544929	0,13581143	0,32176217	0,06451777	0,60708331	0,24805805	0,91489751	0,39053247	0,36409354	0,94249222	0,12812588

Estimating FE regressions

[SA12-Q2b]

Recall from the lecture:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} + \alpha_i + u_{it}$$

We take averages (by i or by t - both are possible):

$$\bar{Y}_i = \beta_0 + \beta_1 \bar{X}_{1,i} + \ldots + \beta_k \bar{X}_{k,i} + \alpha_i + \bar{u}_i$$

We subtract the two equations from each other and get:

$$\left(Y_{it} - \bar{Y}_i\right) = \beta_1 \left(X_{1,it} - \bar{X}_{1,i}\right) + \ldots + \beta_k \left(X_{k,it} - \bar{X}_{k,i}\right) + \left(u_{it} - \bar{u}_i\right)$$

Interpreting FE regressions [SA12-Q2c]

Suppose we estimate this model with fixed effects:

$$\text{Sales}_{it} = \beta_0 + \beta_1 \text{Price}_{it} + \alpha_i + \varepsilon_{it}$$

How do we interpret β_1?

Interpreting FE regressions [SA12-Q2c]

Suppose we estimate this model with fixed effects:

$$\text{Sales}_{it} = \beta_0 + \beta_1 \text{Price}_{it} + \alpha_i + \varepsilon_{it}$$

How do we interpret β_1?

The same way as in an OLS regression!

β_1, is the effect of changes in price on changes in sales, holding constant all other firm-specific factors that do not change over time. In other words, β_1 represents the average effect of a one-unit increase in price on sales within each firm over time, after controlling for the firm-specific fixed effect.

Visualization

See separate gif!

Limitations of fixed effects

- Imagine we want to estimate the effect of annual income on happiness and we have a panel following 1,000 individuals over 10 years. You think there are many omitted variables and include individual-level fixed effects. You are **also** interested in the effect of parental income on happiness. What's the problem?
- We want to estimate the effect of democracy on economic growth in a panel of 180 countries over 60 years. We can include country-fixed effects to account for any effect coming from different countries having a different climate, terrain, access to the sea, culture, etc. Are you satisfied with this approach?

Any questions?

… Remember – Every question is useful!

Section Assignment

See you on R!

Semester Review, Questions, Exam Practice

Econ 140, Section 11

Jonathan Old

Roadmap

1. Recap: Regression Discontinuity

2. Broad topics for the exam

3. Exam practice

Course evaluation



Recap: Regression Discontinuity

The world is full of arbitrary rules

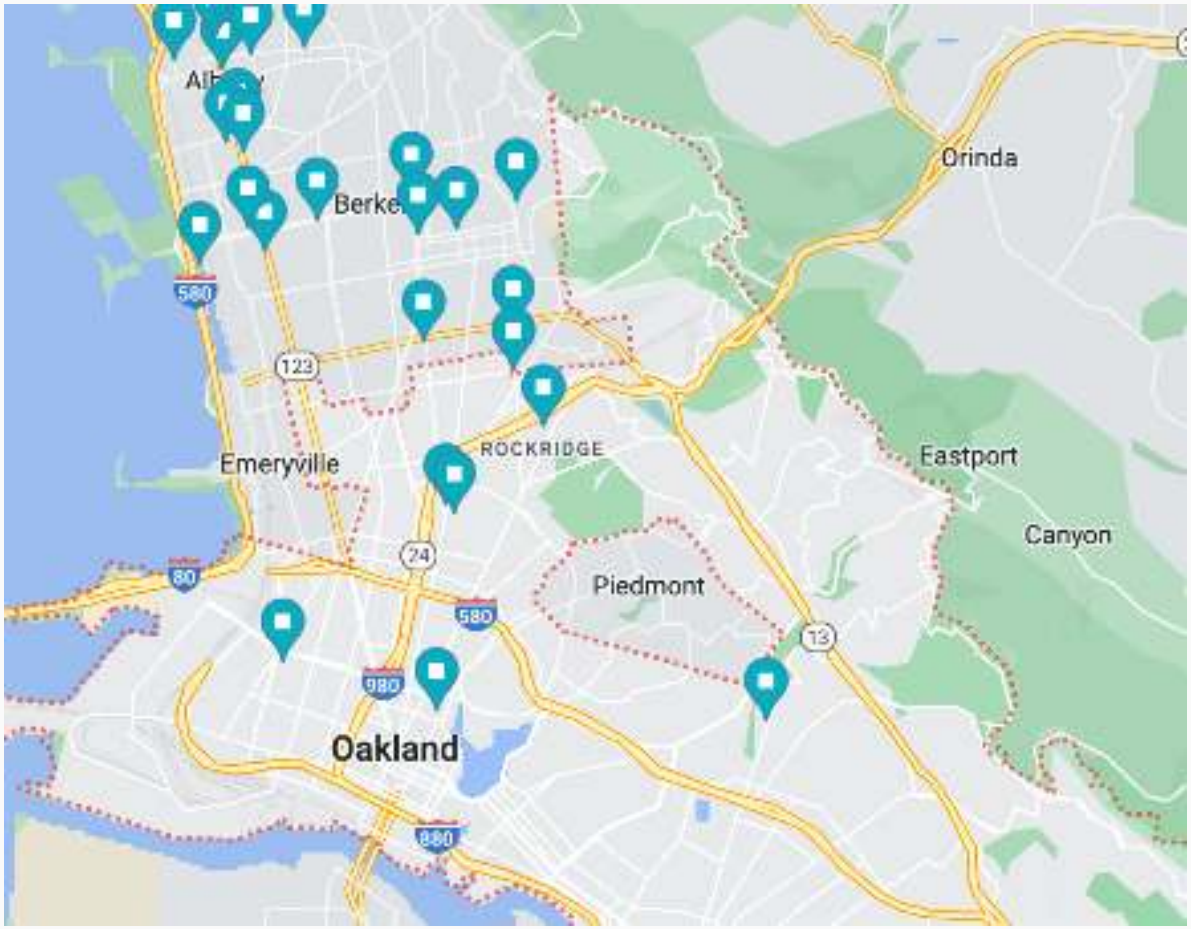
- Students receive a scholarship if their GPA is above 3.0
- Children are allowed to start school if they are five years old before January 31 of that year
- Individuals are eligible for a microfinance loan if they own less than 0.5 acres of land
- Legislators are elected if they receive over 50% of the vote

This creates wonderful discontinuities in the data that we can exploit for econometric analysis!

The most important rules of regression discontinuity

- The essence of RD is to compare people (units) just above the cut-off to others just below
- Crucial assumptions: There are no jumps but our jump at the jump // The effect of X (our running variable) and any other variable C on Y (our outcome) is smooth around the discontinuity
- Identifying assumption: In the absence of the ”treatment” (which was allocated by a discontinuous rule), the outcomes of the treated would have been essentially the same as the outcomes of the untreated
- Limitations:
  1. We only estimate a **local** effect!
  2. There may be strategic behavior at the cutoff (example: school district Oakland/Piedmont)

Oakland and Piedmont



Implementing regression discontinuity designs

With a running variable X, an outcome Y, and a treatment D allocated at the discontinuity, we estimate:

$$Y_i = \alpha + f(X_i) + \beta D_i + \varepsilon_i$$

- $f(X_i)$ is a smooth function of the running variable - e.g. just linear, or quadratic
- We could also estimate whether the **slope** changes at the discontinuity
- We are eventually interested in β: Does the outcome change at the discontinuity?

Broad topics for the exam

Introduction

- Potential outcomes framework
- Interpreting regression tables
- p-values: How surprised we should be to observe the world as it is if your hypothesis about how it works were true?
- Omitted variable bias
- Measurement error
- Reverse causality
- Interpretation of control variables and their role
- Interaction terms (dummy or continuous) – dummyXdummy: get means for every group

RCTs

- Random assignment to treatment solves OVB, reverse causality, measurement error, etc.
- But: Challenges remain
- Spillover effects
- Hawthorne effects
- Imperfect compliance (IV!)
- Non-random attrition

Instrumental variables

- Calculation
- Three assumptions: Relevance, Independence (Exogeneity), Exclusion restriction
- Think of three regressions: First stage, reduced form, ”target regression” (Second stage)
- Connection to RCTs (imperfect compliance)
- Local Average Treatment Effect (ITT/TOT/...)
- Connect with potential outcomes framework

Differences in Differences

- Connect with potential outcomes
- Parallel trends assumption
- Calculate with tables
- Calculate using regression tables

Panel Data and Fixed effects

- Fixed effects: Calculation
- First-differences
- Useful way to think about it: We add a lot of controls, but key problem remains (time-varying confounders, ...)
- Connection of two-way fixed effects with DiD

Regression Discontinuities

- ”No jump but our jump at the jump”
- Fuzzy DiD is just IV

General and overarching concepts

- Potential outcomes: Please review carefully – maybe not on the long questions but who knows?
- Comparison of OLS and other estimates: What changed and why? (Hint: Always go back to OVB, less likely measurement error or reverse causality)
- Be creative but also clear: Most thinking questions can be solved with reasonable economic intuition, your knowledge of the world, and what you learned on the course.
- Keep in mind the big picture: We are just trying to find a valid counterfactual, to estimate causal effects!

Exam practice

RDD question

Question 2

The local government wants to estimate the impact on future earnings of a job training program that it operated in 2016 and 2017. Access to the program is governed by an eligibility rule: only individuals whose income in the prior tax year was less than £12,000 can participate.

A. **(20 marks)** Explain how you could use this eligibility rule to estimate the causal effect of the program. Describe any data you would need, write down the regression equation(s) you would estimate, define all variables precisely, and explain how you would interpret the regression results. If you make any additional assumptions, state them clearly.

B. **(10 marks)** A colleague worries that because the eligibility rule was public, your estimates of the program's causal effect may be biased. Why might this pose a problem for identification? How could you use the data to assess the validity of this concern?

DiD question

Question 3

(10 marks) On 28 August 2017, a new state law forced the city of St. Louis, Missouri, United States to reduce its minimum wage from $10 per hour to $7.70 per hour. You have monthly employment data for at the municipal level for the entire state in the months of May 2017 and November 2017. How would you construct the difference-in-differences estimate for the effect of reducing the minimum wage? What assumption would have to hold for this to be a valid causal estimate of the effect? Precisely describe one concern that you may have with the validity of this assumption. Describe two things you could do to assess the validity of this assumption.