Week 11: Causality with Unmeasured Confounding

Brandon Stewart[1]

Princeton

December 3 and 5, 2018

[1] These slides are heavily influenced by Matt Blackwell, Adam Glynn, Jens Hainmueller

Where We've Been and Where We're Going...

Where We've Been and Where We're Going...

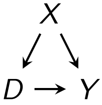
- Last Week
  - selection on observables and measured confounding
- This Week
  - Monday:
    - natural experiments
    - classical view of instrumental variables
  - Wednesday:
    - modern view of instrumental variables
    - regression discontinuity
- The Following Week
  - repeated observations and wrap up
- Long Run
  - probability → inference → regression → causal inference

Questions?

Unmeasured Confounding

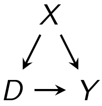
Unmeasured Confounding

- Last week we considered cases of measured confounding



Unmeasured Confounding

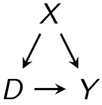
- Last week we considered cases of measured confounding



- In this case we block the backdoor path $D \leftarrow X \rightarrow Y$ by conditioning on X.

Unmeasured Confounding

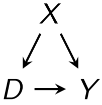
- Last week we considered cases of measured confounding



- In this case we block the backdoor path $D \leftarrow X \rightarrow Y$ by conditioning on X.
- What happens in the general case where X is unobserved?

Unmeasured Confounding

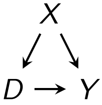
- Last week we considered cases of measured confounding



- In this case we block the backdoor path $D \leftarrow X \rightarrow Y$ by conditioning on X.
- What happens in the general case where X is unobserved?
- Under selection on unobservables we are going to need a different approach which we will talk about over the next two weeks.

Unmeasured Confounding

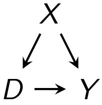
- Last week we considered cases of measured confounding



- In this case we block the backdoor path $D \leftarrow X \rightarrow Y$ by conditioning on X.
- What happens in the general case where X is unobserved?
- Under selection on unobservables we are going to need a different approach which we will talk about over the next two weeks.
- There is No Free Lunch $\rightsquigarrow$ we can't get something for nothing, we will need new variables, new assumptions and new approaches.

Unmeasured Confounding

- Last week we considered cases of measured confounding



- In this case we block the backdoor path $D \leftarrow X \rightarrow Y$ by conditioning on X.
- What happens in the general case where X is unobserved?
- Under selection on unobservables we are going to need a different approach which we will talk about over the next two weeks.
- There is No Free Lunch $\rightsquigarrow$ we can't get something for nothing, we will need new variables, new assumptions and new approaches.
- Goal: give you a feel for what is possible, but note that you will need to do work beyond class if you want to use one of these techniques.

Approaches to Unmeasured Confounding

- Natural Experiments

Approaches to Unmeasured Confounding

- Natural Experiments
- Interrupted Time-Series

Approaches to Unmeasured Confounding

- Natural Experiments
- Interrupted Time-Series
- Instrumental Variables

Approaches to Unmeasured Confounding

- Natural Experiments
- Interrupted Time-Series
- Instrumental Variables
- Regression Discontinuity

Approaches to Unmeasured Confounding

- Natural Experiments
- Interrupted Time-Series
- Instrumental Variables
- Regression Discontinuity
- Bounding

Approaches to Unmeasured Confounding

- Natural Experiments
- Interrupted Time-Series
- Instrumental Variables
- Regression Discontinuity
- Bounding
- Sensitivity Analysis

Approaches to Unmeasured Confounding

- Natural Experiments
- Interrupted Time-Series
- Instrumental Variables
- Regression Discontinuity
- Bounding
- Sensitivity Analysis
- Front Door Adjustment

Approaches to Unmeasured Confounding

- Natural Experiments
- Interrupted Time-Series
- Instrumental Variables
- Regression Discontinuity
- Bounding
- Sensitivity Analysis
- Front Door Adjustment

Approaches to Unmeasured Confounding

- Natural Experiments (today)
- Interrupted Time-Series (today)
- Instrumental Variables (today and Wednesday)
- Regression Discontinuity (Wednesday)
- Bounding
- Sensitivity Analysis
- Front Door Adjustment

Natural Experiments

Natural Experiments

- Broadly speaking an "experiment" where the treatment is randomized but the randomization was not controlled by the researcher.

Natural Experiments

- Broadly speaking an "experiment" where the treatment is randomized but the randomization was not controlled by the researcher.
- Leverages an exogenous (external to the system) event to measure the effect of an otherwise endogenous (internal to the system) phenomenon.

Natural Experiments

- Broadly speaking an "experiment" where the treatment is randomized but the randomization was not controlled by the researcher.
- Leverages an exogenous (external to the system) event to measure the effect of an otherwise endogenous (internal to the system) phenomenon.
- Trickier to analyze than regular experiments

Natural Experiments

- Broadly speaking an "experiment" where the treatment is randomized but the randomization was not controlled by the researcher.
- Leverages an exogenous (external to the system) event to measure the effect of an otherwise endogenous (internal to the system) phenomenon.
- Trickier to analyze than regular experiments
  - we ought to be suspicious of randomization we don't control

Natural Experiments

- Broadly speaking an "experiment" where the treatment is randomized but the randomization was not controlled by the researcher.
- Leverages an exogenous (external to the system) event to measure the effect of an otherwise endogenous (internal to the system) phenomenon.
- Trickier to analyze than regular experiments
  - we ought to be suspicious of randomization we don't control
  - nature may not choose exactly the treatment we want

Natural Experiments

- Broadly speaking an "experiment" where the treatment is randomized but the randomization was not controlled by the researcher.
- Leverages an exogenous (external to the system) event to measure the effect of an otherwise endogenous (internal to the system) phenomenon.
- Trickier to analyze than regular experiments
  - we ought to be suspicious of randomization we don't control
  - nature may not choose exactly the treatment we want
  - not immediately obvious which groups are comparable

Natural Experiments

- Broadly speaking an "experiment" where the treatment is randomized but the randomization was not controlled by the researcher.
- Leverages an exogenous (external to the system) event to measure the effect of an otherwise endogenous (internal to the system) phenomenon.
- Trickier to analyze than regular experiments
  - we ought to be suspicious of randomization we don't control
  - nature may not choose exactly the treatment we want
  - not immediately obvious which groups are comparable
  - valid comparison may not estimate the causal effect of interest

Natural Experiments

- Broadly speaking an "experiment" where the treatment is randomized but the randomization was not controlled by the researcher.
- Leverages an exogenous (external to the system) event to measure the effect of an otherwise endogenous (internal to the system) phenomenon.
- Trickier to analyze than regular experiments
  - we ought to be suspicious of randomization we don't control
  - nature may not choose exactly the treatment we want
  - not immediately obvious which groups are comparable
  - valid comparison may not estimate the causal effect of interest
- When available, a useful way to capitalize on randomness in the world to make causal inferences.

Natural Experiments

- Broadly speaking an "experiment" where the treatment is randomized but the randomization was not controlled by the researcher.
- Leverages an exogenous (external to the system) event to measure the effect of an otherwise endogenous (internal to the system) phenomenon.
- Trickier to analyze than regular experiments
  - we ought to be suspicious of randomization we don't control
  - nature may not choose exactly the treatment we want
  - not immediately obvious which groups are comparable
  - valid comparison may not estimate the causal effect of interest
- When available, a useful way to capitalize on randomness in the world to make causal inferences.
- See Dunning (2012) *Natural Experiments in the Social Sciences*

Caution on terminology

It is worth nothing that the label "natural experiment" is perhaps unfortunate. As we shall see, the social and political forces that give rise to as-if random assignment of interventions are not generally "natural" in the ordinary sense of that term. Second, natural experiments are observational studies, not true experiments, again, because they lack an experimental manipulation. In sum, natural experiments are neither natural nor experiments.

—Dunning (2012) pg 16

Natural Experiment Examples (True Randomization)

Natural Experiment Examples (True Randomization)

Randomness	**Focus**	**Citation**
Vietnam draft	labor market	Angrist 1990
randomized quotas	female leadership in Indian village council presidencies	Chattopadhyay & Duflo 2004
randomized term lengths	tenure in office on legislative performance	Dal Bo & Rossi 2010
lottery	effect of winnings on political attitudes	Doherty, Green & Gerber 2006
randomized ballot order	ballot order effects in CA	Ho & Imai 2008

Natural Experiment Examples (As If Randomization)

Randomness	Focus	Citation
child abduction by LRA	child soldering affecting political participation	Blattman 2008
election monitor assignment	international election monitoring on fraud	Hyde 2007
random shelling by drunk soldiers	indiscriminate violence on rebellion	Lyall 2009
hurricane	study of friendship formulation	Phan and Airoldi 2015
2006 Israel-Hezbollah war	stress on unborn babies	Torche and Shwed 2015
Snowden revelations	reading behavior on wikipedia	Penney 2016
terrorist attacks	perception of immigrants	Legewie 2013

Questions to Ask Yourself

Questions to Ask Yourself

From Sekhon and Titiunik (2012) "When Natural Experiments Are Neither Natural nor Experiments" *American Political Science Review*

1. "is the proposed treatment-control comparison guaranteed to be valid by the assumed randomization?"

Questions to Ask Yourself

From Sekhon and Titiunik (2012) "When Natural Experiments Are Neither Natural nor Experiments" *American Political Science Review*

1. "is the proposed treatment-control comparison guaranteed to be valid by the assumed randomization?"
2. "if not, what is the comparison that is guaranteed by the randomization, and how does this comparison relate to the comparison the researcher wishes to make?

Example: Redistricting

Sekhon and Titiunik 2012 discussion of Ansolabehere et al. 2000

Example: Redistricting

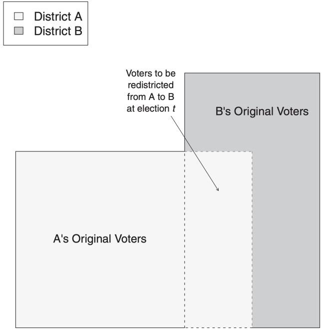
Sekhon and Titiunik 2012 discussion of Ansolabehere et al. 2000

- problem: difficult to estimate incumbency advantage.

Example: Redistricting

Sekhon and Titiunik 2012 discussion of Ansolabehere et al. 2000

- problem: difficult to estimate incumbency advantage.
- proposed solution: redistricting as a natural experiment

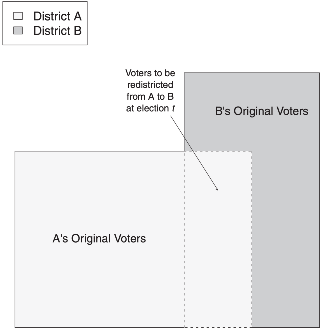


(a) Before One-time Redistricting

Example: Redistricting

Sekhon and Titiunik 2012 discussion of Ansolabehere et al. 2000

- problem: difficult to estimate incumbency advantage.
- proposed solution: redistricting as a natural experiment
- two types of variation: temporal (voting before and after) and cross-sectional (some voters move)

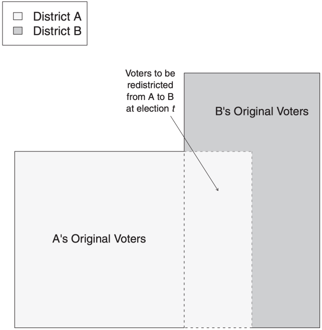


(a) Before One-time Redistricting

Example: Redistricting

Sekhon and Titiunik 2012 discussion of Ansolabehere et al. 2000

- problem: difficult to estimate incumbency advantage.
- proposed solution: redistricting as a natural experiment
- two types of variation: temporal (voting before and after) and cross-sectional (some voters move)
- key claim: if district is randomly redrawn, we can attribute voting differences between new voters in a district and their new neighbors to a lack of incumbency advantage

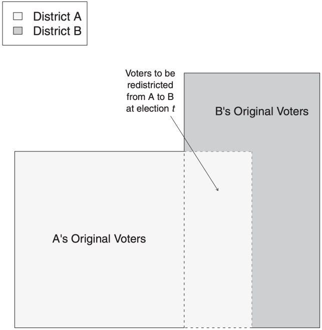


(a) Before One-time Redistricting

Example: Redistricting

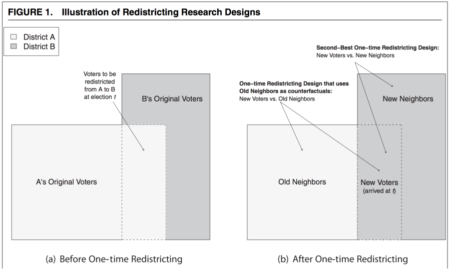
Sekhon and Titiunik 2012 discussion of Ansolabehere et al. 2000

- problem: difficult to estimate incumbency advantage.
- proposed solution: redistricting as a natural experiment
- two types of variation: temporal (voting before and after) and cross-sectional (some voters move)
- key claim: if district is randomly redrawn, we can attribute voting differences between new voters in a district and their new neighbors to a lack of incumbency advantage
- idea is that two groups have same incumbent, same challenger, same campaign environment, but different histories with incumbent



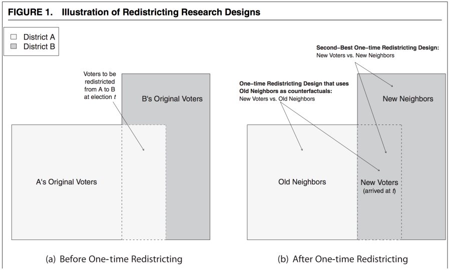
(a) Before One-time Redistricting

Example: Redistricting



1) Is the comparison of B's old voters and B's new voters guaranteed to be valid if we assume that voters are redistricted randomly?

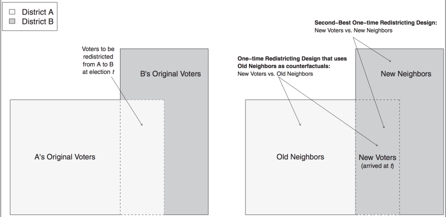
Example: Redistricting



1) Is the comparison of B's old voters and B's new voters guaranteed to be valid if we assume that voters are redistricted randomly?
   - No! only voters in A were subject to randomization

Example: Redistricting

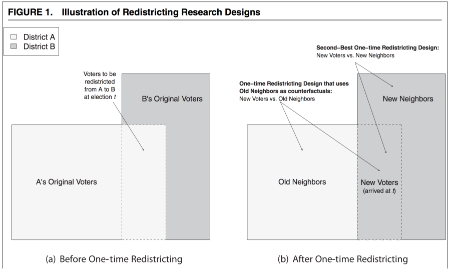
FIGURE 1. Illustration of Redistricting Research Designs



(a) Before One-time Redistricting (b) After One-time Redistricting

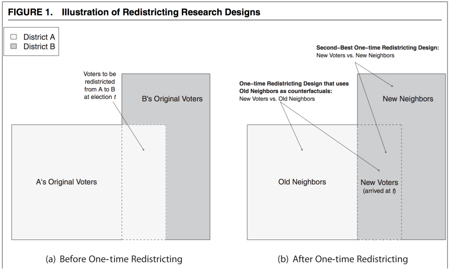
1) Is the comparison of B's old voters and B's new voters guaranteed to be valid if we assume that voters are redistricted randomly?
   - No! only voters in A were subject to randomization
   - Natural experiment creates three distinct groups: new voters, new neighbors, old neighbors and not all comparisons are valid.

Example: Redistricting



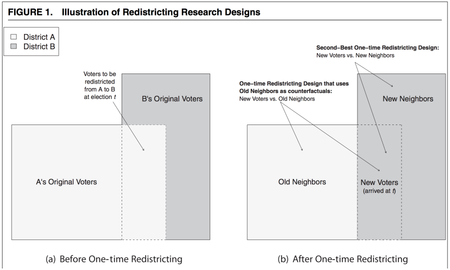
1) Is the comparison of B's old voters and B's new voters guaranteed to be valid if we assume that voters are redistricted randomly?
   - No! only voters in A were subject to randomization
   - Natural experiment creates three distinct groups: new voters, new neighbors, old neighbors and not all comparisons are valid.
2) What comparison is guaranteed to be valid if redistricting is done at random?

Example: Redistricting



1) Is the comparison of B's old voters and B's new voters guaranteed to be valid if we assume that voters are redistricted randomly?
   - No! only voters in A were subject to randomization
   - Natural experiment creates three distinct groups: new voters, new neighbors, old neighbors and not all comparisons are valid.
2) What comparison is guaranteed to be valid if redistricting is done at random?
   - random redistricting guarantees that old neighbors and new voters are comparable.

Example: Redistricting



1) Is the comparison of B's old voters and B's new voters guaranteed to be valid if we assume that voters are redistricted randomly?
    - No! only voters in A were subject to randomization
    - Natural experiment creates three distinct groups: new voters, new neighbors, old neighbors and not all comparisons are valid.
2) What comparison is guaranteed to be valid if redistricting is done at random?
    - random redistricting guarantees that old neighbors and new voters are comparable.
    - need to find a new design (see Sekhon and Titiunik 2012 for more)

Pitfalls

- We still really need theory to guide our thinking

Pitfalls

- We still really need theory to guide our thinking
- Understanding the assignment and causal process is extremely important (was this really random? is the treatment really what we care about?)

Pitfalls

- We still really need theory to guide our thinking
- Understanding the assignment and causal process is extremely important (was this really random? is the treatment really what we care about?)
- The result only applies to a limited population — is it significant?

Pitfalls

- We still really need theory to guide our thinking
- Understanding the assignment and causal process is extremely important (was this really random? is the treatment really what we care about?)
- The result only applies to a limited population — is it significant?
- Be sure to verify that your "as-if-random" assignment is really random (e.g. placebo tests, balance tests)

Pitfalls

- We still really need theory to guide our thinking
- Understanding the assignment and causal process is extremely important (was this really random? is the treatment really what we care about?)
- The result only applies to a limited population — is it significant?
- Be sure to verify that your "as-if-random" assignment is really random (e.g. placebo tests, balance tests)
- If possible, use sensitivity tests to evaluate susceptibility to unobserved confounding (e.g. Cinelli and Hazlett 2018)

Pitfalls

- We still really need theory to guide our thinking
- Understanding the assignment and causal process is extremely important (was this really random? is the treatment really what we care about?)
- The result only applies to a limited population — is it significant?
- Be sure to verify that your "as-if-random" assignment is really random (e.g. placebo tests, balance tests)
- If possible, use sensitivity tests to evaluate susceptibility to unobserved confounding (e.g. Cinelli and Hazlett 2018)
- Convincingly analyzing a natural experiment takes at least as much careful thought not less!

Reasons to Be Excited

Reasons to Be Excited

- Now that you know what to look for you may see more natural experiments out there

Reasons to Be Excited

- Now that you know what to look for you may see more natural experiments out there
- Exogenous randomization can help us make credible causal inferences in places where we never could have run an experiment

Reasons to Be Excited

- Now that you know what to look for you may see more natural experiments out there
- Exogenous randomization can help us make credible causal inferences in places where we never could have run an experiment
- It is often pretty easy to communicate these kinds of methods to non-experts

Reasons to Be Excited

- Now that you know what to look for you may see more natural experiments out there
- Exogenous randomization can help us make credible causal inferences in places where we never could have run an experiment
- It is often pretty easy to communicate these kinds of methods to non-experts
- Salganik (2017) argues that with always-on digital data collection we will be in better shape moving forward to leverage natural experiments as the opportunities arise.

Interrupted Time Series

Interrupted Time Series

- A simple construction often used with natural experiment is the Interrupted Time Series(ITS)

Interrupted Time Series

- A simple construction often used with natural experiment is the Interrupted Time Series(ITS)
- ITS designs convey the basic intuition that when an event abruptly occurs, we can compare results immediately before and immediately afterwards.

Interrupted Time Series

- A simple construction often used with natural experiment is the Interrupted Time Series(ITS)
- ITS designs convey the basic intuition that when an event abruptly occurs, we can compare results immediately before and immediately afterwards.
- We can write this as a model:

$$Y_t = f(t) + D_t\beta + \epsilon_t$$

Interrupted Time Series

- A simple construction often used with natural experiment is the Interrupted Time Series(ITS)
- ITS designs convey the basic intuition that when an event abruptly occurs, we can compare results immediately before and immediately afterwards.
- We can write this as a model:

$$Y_t = f(t) + D_t\beta + \epsilon_t$$

- The key identifying assumption is that the observed values of y_t before the treatment status switches at t^* can be used to specify $f(t)$ for the rest of the series used.

Interrupted Time Series Example

American Political Science Review (2018) 112, 3, 621–636
doi:10.1017/S0003055418000084

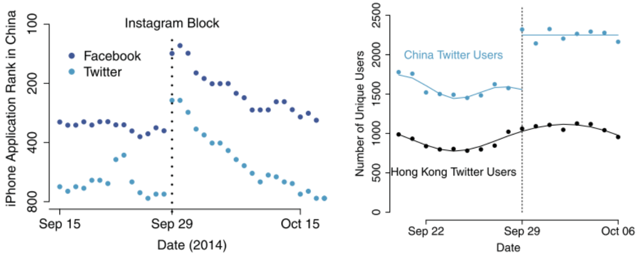
How Sudden Censorship Can Increase Access to Information

WILLIAM R. HOBBS *Northeastern University*
MARGARET E. ROBERTS *University of California, San Diego*

Conventional wisdom assumes that increased censorship will strictly decrease access to information. We delineate circumstances when increases in censorship expand access to information for a substantial subset of the population. When governments suddenly impose censorship on previously uncensored information, citizens accustomed to acquiring this information will be incentivized to learn methods of censorship evasion. These evasion tools provide continued access to the newly blocked information—and also extend users' ability to access information that has long been censored. We illustrate this phenomenon using millions of individual-level actions of social media users in China before and after the block of Instagram. We show that the block inspired millions of Chinese users to acquire virtual private networks, and that these users subsequently joined censored websites like Twitter and Facebook. Despite initially being apolitical, these new users began browsing blocked political pages on Wikipedia, following Chinese political activists on Twitter, and discussing highly politicized topics such as opposition protests in Hong Kong.

Interrupted Time Series Example

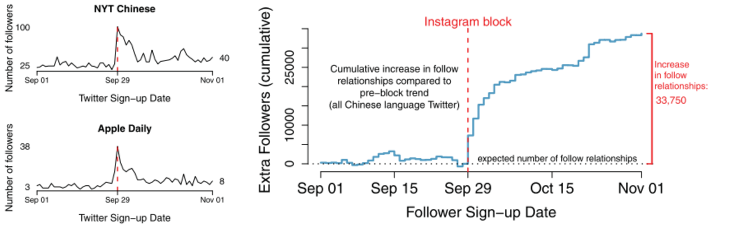
FIGURE 3. Left: The Instagram block's effect on the rank of Facebook and Twitter on iPhones from mainland China, from AppAnnie.com. Right: Comparison of tweets per day from Mainland China and Hong Kong before and after the Instagram block.



The left panel of this figure shows the change in download ranks for Facebook and Twitter before and after Instagram was blocked. The right panel of this figure shows that the Chinese Twitter users in our sample increased 30% the same day that we observe a spike in Instagram mentions and several days after the beginning of the Hong Kong protests. This increase only occurred in China and not in Hong Kong. The lines in this panel were fit using a smoothing spline.

Interrupted Time Series Example

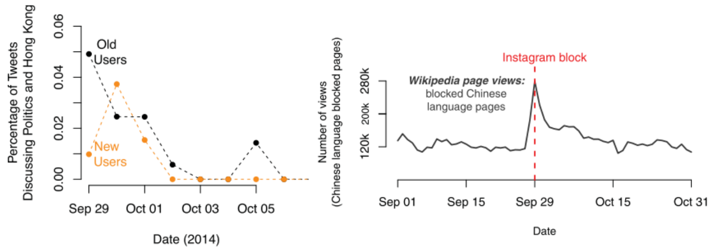
FIGURE 4. Left: Daily new followers to *New York Times* Chinese and *Apple Daily* Twitter accounts (based on new user sign-up dates). Right: Cumulative increase in followers, compared to preblock trend, of any Chinese language user (based on new user sign-up dates) compared to expected increase in followers.



The left panel of this figure shows the sign-up dates of followers of the *New York Times* Chinese and *Apple Daily* Twitter accounts. Many followers of these accounts signed up for Twitter immediately following the Instagram block. This increase in sign-ups—users who eventually followed *NYT* Chinese and *Apple Daily*—continues long after the Instagram block. The right panel of this figure shows that all Chinese language Twitter users accumulated approximately 33,750 more followers from new Twitter sign-ups than what we would expect based on pre-block trends. This cumulative increase was calculated using a cumulative sum of the number of new followers minus the number of expected followers, where the expected followers was the mean daily number of new followers prior to the Instagram block.

Interrupted Time Series Example

FIGURE 5. Left: Tweets that mention politics in Hong Kong, comparison of new users and old users. Right: Page views for Chinese language Wikipedia pages blocked in China. Bottom: Changes in Wikipedia views.



Largest increases in views September 28th to 29th *First day of Instagram block*	**Largest increases in views September 29th to 30th** *Second day of Instagram block*
Occupy Central	**People's Republic of China blocked websites list**
Hong Kong	**Jiang Zemin**
June 4th Incident	Radio Australia
Hong Kong Independence Movement	**Hu Jintao**
pan-Democrats	**Zeng Qing**
Xi Jinping	**Wang Weilin (Tank Man)**
Sunflower student movement	**Li Peng**
Guo Baixiong	**Ling Jihua**
Article 23 Hong Kong	**Tiananmen Square Incident**
Zhang Xuan (celebrity who supported HK movement)	**Zhou Yongkang**
Alliance for True Democracy	**Wu'erkaixi (June 4 leader)**
Bo News (dissident Chinese news network, located in U.S.)	**Zhang Dejiang**
Chai Ling (June 4 leader)	YouTube
Jasmine Revolution	**Wen Jiabao**
Hu Yaobang	AV Actress (Japanese porn stars)
Taiwan	**Mao Zedong**
Jiang Zemin	**Deng Xiaoping**
People's Republic of China blocked websites list	**Ling Gu (son of Ling Jihua)**
Democracy	**Wu Bangguo**
Wu'erkaixi (June 4 leader)	**Hua Guofeng**

Motivating Instrumental Variables

Motivating Instrumental Variables

- We saw how to identify and estimate effects under no unmeasured confounding and just now when nature happens to randomize the treatment we want.

Motivating Instrumental Variables

- We saw how to identify and estimate effects under no unmeasured confounding and just now when nature happens to randomize the treatment we want.
- Are we doomed if neither happens?

Motivating Instrumental Variables

- We saw how to identify and estimate effects under no unmeasured confounding and just now when nature happens to randomize the treatment we want.
- Are we doomed if neither happens?
- Instrumental variables (IV) allow us to exploit an exogenous source of variation that drives the treatment but does not otherwise affect the outcome.

Motivating Instrumental Variables

- We saw how to identify and estimate effects under no unmeasured confounding and just now when nature happens to randomize the treatment we want.
- Are we doomed if neither happens?
- Instrumental variables (IV) allow us to exploit an exogenous source of variation that drives the treatment but does not otherwise affect the outcome.
- If we have an instrument, we can deal with unmeasured confounding in the treatment-outcome relationship.

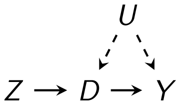
Motivating Instrumental Variables

- We saw how to identify and estimate effects under no unmeasured confounding and just now when nature happens to randomize the treatment we want.
- Are we doomed if neither happens?
- Instrumental variables (IV) allow us to exploit an exogenous source of variation that drives the treatment but does not otherwise affect the outcome.
- If we have an instrument, we can deal with unmeasured confounding in the treatment-outcome relationship.
- It is going to turn out that the same construction will let us deal with non-compliance in experiments.

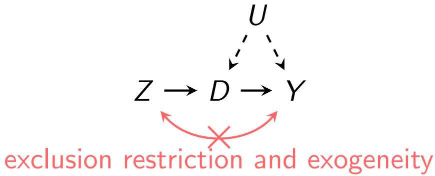
Angrist (1990): Draft lottery as an instrument to study the relationship between military service and income

https://en.wikipedia.org/wiki/Draft_lottery_(1969)#/media/File:1969_draft_lottery_photo.jpg

Graphical Model

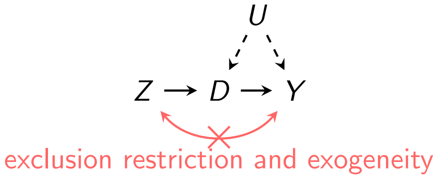


Graphical Model



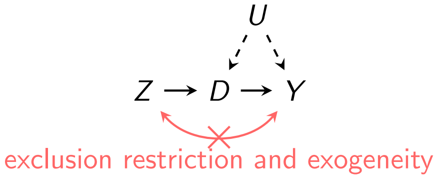
- Notation: Z is the instrument, D is the treatment, and U is the unmeasured confounder

Graphical Model



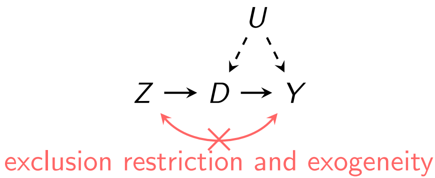
- Notation: Z is the instrument, D is the treatment, and U is the unmeasured confounder
- (Approximate) assumptions (3 of 4)
  1) instrument/treatment and instrument/outcome don't share unmeasured common causes (exogeneity of the instrument)
  2) no direct or indirect effect of the instrument on the outcome not through the treatment (exclusion restriction)

Graphical Model



- Notation: Z is the instrument, D is the treatment, and U is the unmeasured confounder
- (Approximate) assumptions (3 of 4)
  1) instrument/treatment and instrument/outcome don't share unmeasured common causes (exogeneity of the instrument)
  2) no direct or indirect effect of the instrument on the outcome not through the treatment (exclusion restriction)
  3) Z affects D (first stage relationship)

Graphical Model



- Notation: Z is the instrument, D is the treatment, and U is the unmeasured confounder
- (Approximate) assumptions (3 of 4)
  1) instrument/treatment and instrument/outcome don't share unmeasured common causes (exogeneity of the instrument)
  2) no direct or indirect effect of the instrument on the outcome not through the treatment (exclusion restriction)
  3) Z affects D (first stage relationship)
- We will need one more later which we will come back to.

Some Examples

- Angrist (1990): Draft lottery as an IV for military service (income as outcome)
- Acemoglu et al (2001): settler mortality as an IV for institutional quality (GDP/capita as outcome)
- Levitt (1997): being an election year as IV for police force size (crime as outcome)
- Miguel, Satayanath & Sergenti (2004): lagged rainfall as IV for GDP per capita effect (outcome is civil war onset).
- Kern & Hainmueller (2009): having West German TV reception in East Berlin as an instrument for West German TV watching (outcome is support for the East German regime)
- Nunn & Wantchekon (2011): historical distance of ethnic group to the coast as a instrument for the slave raiding of that ethnic group (outcome are trust attitudes today)

Core Idea

Core Idea

The world has randomized something (the instrument) just maybe not the thing you want (the treatment).

Core Idea

The world has randomized something (the instrument) just maybe not the thing you want (the treatment).

Subject to four assumptions you may be able to get (approximately) what you want anyway.

Non-Compliance Motivation for Instrumental Variables

Non-Compliance Motivation for Instrumental Variables

Problem

Non-Compliance Motivation for Instrumental Variables

Problem

- *Often we cannot force subjects to take specific treatments*

Non-Compliance Motivation for Instrumental Variables

Problem

- *Often we cannot force subjects to take specific treatments*
- *Units choosing to take the treatment may differ in unobserved characteristics from units that refrain from doing so*

Non-Compliance Motivation for Instrumental Variables

Problem

- *Often we cannot force subjects to take specific treatments*
- *Units choosing to take the treatment may differ in unobserved characteristics from units that refrain from doing so*

Example: Non-compliance in JTPA Experiment

	Not Enrolled in Training	Enrolled in Training	Total
Assigned to Control	3,663	54	3,717
Assigned to Training	2,683	4,804	7,487
Total	6,346	4,858	11,204

Two Views on Instrumental Variables

Two Views on Instrumental Variables

1. Traditional Econometric Framework
   - strong assumptions
     - ★ constant effects
     - ★ linearity in case of a continuous treatment
   - Identifies the average treatment effect

Two Views on Instrumental Variables

1. Traditional Econometric Framework
   - strong assumptions
     - constant effects
     - linearity in case of a continuous treatment
   - Identifies the average treatment effect
2. Potential Outcome Model of IV
   - Weaker assumptions
     - monotonicity
     - allows heterogeneous treatment effect
   - Only identifies Local Average Treatment Effect (LATE)

The Problem (informal)

The Problem (informal)

Suppose want to know the average effect of X on Y.

The Problem (informal)

Suppose want to know the average effect of X on Y.

Two Problems:

The Problem (informal)

Suppose want to know the average effect of X on Y.

Two Problems:

1. We may not be able to measure all variables that affect both X and Y.

The Problem (informal)

Suppose want to know the average effect of X on Y.

Two Problems:

1. We may not be able to measure all variables that affect both X and Y.
2. We may not be able to measure X without error.

The Problem (informal)

Suppose want to know the average effect of X on Y.

Two Problems:

1. We may not be able to measure all variables that affect both X and Y.
2. We may not be able to measure X without error.

Both of these conditions will induce bias in our effect estimates.

The Problem (formal)

The Problem (formal)

Assume a linear structural equation model but suppose that the classical "exogeneity" condition ($E[U_i|X_i] = 0$) does not hold:

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

The Problem (formal)

Assume a linear structural equation model but suppose that the classical "exogeneity" condition ($E[U_i|X_i] = 0$) does not hold:

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

This can happen for a number of reasons including:

- omitted variables (which is a common cause of X and Y)
- measurement error in X
- included variables (post-treatment or M-structures)
- simultaneous equations (endogenous feedback loops)

The Problem (formal)

Assume a linear structural equation model but suppose that the classical "exogeneity" condition ($E[U_i|X_i] = 0$) does not hold:

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

This can happen for a number of reasons including:

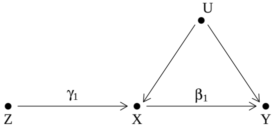
- omitted variables (which is a common cause of X and Y)
- measurement error in X
- included variables (post-treatment or M-structures)
- simultaneous equations (endogenous feedback loops)

We will typically formulate the problem as resulting from omitted confounding.

A Potential Solution: Instrumental Variables (IV)

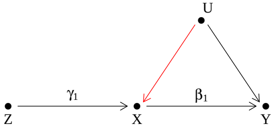
$$Y_i = \beta_0 + \beta_1 X_i + U_i$$

$$X_i = \gamma_0 + \gamma_1 Z_i + U_i$$



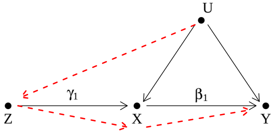
A Potential Solution: Instrumental Variables (IV)

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$
$$E[U_i|X_i] \neq 0$$
$$X_i = \gamma_0 + \gamma_1 Z_i + U_i$$



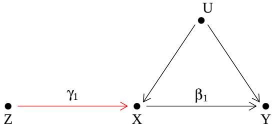
A Potential Solution: Instrumental Variables (IV)

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$
$$E[U_i|X_i] \neq 0$$
$$X_i = \gamma_0 + \gamma_1 Z_i + U_i$$
$$E[U_i|Z_i] = 0$$



A Potential Solution: Instrumental Variables (IV)

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_i + U_i \\
&E[U_i|X_i] \neq 0 \\
X_i &= \gamma_0 + \gamma_1 Z_i + U_i \\
&E[U_i|Z_i] = 0 \\
&Cov[X_i, Z_i] \neq 0
\end{aligned}
$$



Commonly Used Instrumental Variables

- Assigned status in randomized trials with noncompliance

Commonly Used Instrumental Variables

- Assigned status in randomized trials with noncompliance
  - assigned vs. enrolled in job training

Commonly Used Instrumental Variables

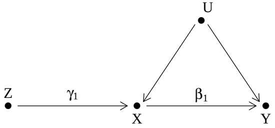
- Assigned status in randomized trials with noncompliance
  - assigned vs. enrolled in job training
  - "received" versus "read" the mailing in the social pressure experiment

Commonly Used Instrumental Variables

- Assigned status in randomized trials with noncompliance
  - assigned vs. enrolled in job training
  - "received" versus "read" the mailing in the social pressure experiment
- Rainfall, earthquakes, ...

Commonly Used Instrumental Variables

- Assigned status in randomized trials with noncompliance
  - assigned vs. enrolled in job training
  - "received" versus "read" the mailing in the social pressure experiment
- Rainfall, earthquakes, ...
- ...



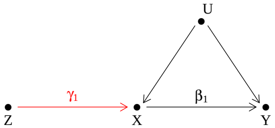
The IV Estimator

With our assumed model,

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$
$$X_i = \gamma_0 + \gamma_1 Z_i + U_i$$

Assuming the model is true,

- regressing X on Z consistently estimates γ_1



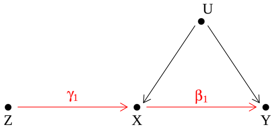
The IV Estimator

With our assumed model,

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$
$$X_i = \gamma_0 + \gamma_1 Z_i + U_i$$

Assuming the model is true,

- regressing X on Z consistently estimates γ_1
- regressing Y on Z consistently estimates $\gamma_1 \cdot \beta_1$



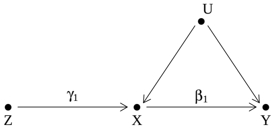
The IV Estimator

With our assumed model,

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$
$$X_i = \gamma_0 + \gamma_1 Z_i + U_i$$

Assuming the model is true,

- regressing X on Z consistently estimates γ_1
- regressing Y on Z consistently estimates $\gamma_1 \cdot \beta_1$
- $\frac{\widehat{\gamma_1 \cdot \beta_1}}{\widehat{\gamma_1}}$ consistently estimates $\frac{\gamma_1 \cdot \beta_1}{\gamma_1} = \beta_1$



Preview: The Problem of Weak Instruments

Preview: The Problem of Weak Instruments

Notice that the IV technique uses a ratio $\frac{\gamma_1 \cdot \beta_1}{\gamma_1} = \beta_1$.

Preview: The Problem of Weak Instruments

Notice that the IV technique uses a ratio $\frac{\gamma_1 \cdot \beta_1}{\gamma_1} = \beta_1$.

Dividing by zero (or near zero) makes things blow up.

Preview: The Problem of Weak Instruments

Notice that the IV technique uses a ratio $\frac{\gamma_1 \cdot \beta_1}{\gamma_1} = \beta_1$.

Dividing by zero (or near zero) makes things blow up.

Therefore, if the instrument is weak ($\gamma_1 \approx 0$), and our estimates of γ_1 and $\gamma_1 \cdot \beta_1$ are not perfect, we can get inaccurate estimates of β_1:

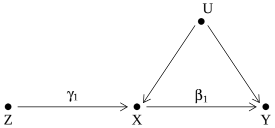
Preview: The Problem of Weak Instruments

Notice that the IV technique uses a ratio $\frac{\gamma_1 \cdot \beta_1}{\gamma_1} = \beta_1$.

Dividing by zero (or near zero) makes things blow up.

Therefore, if the instrument is weak ($\gamma_1 \approx 0$), and our estimates of γ_1 and $\gamma_1 \cdot \beta_1$ are not perfect, we can get inaccurate estimates of β_1:

- medium sample size $\Rightarrow$ high variance



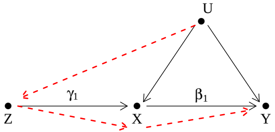
Preview: The Problem of Weak Instruments

Notice that the IV technique uses a ratio $\frac{\gamma_1 \cdot \beta_1}{\gamma_1} = \beta_1$.

Dividing by zero (or near zero) makes things blow up.

Therefore, if the instrument is weak ($\gamma_1 \approx 0$), and our estimates of γ_1 and $\gamma_1 \cdot \beta_1$ are not perfect, we can get inaccurate estimates of β_1:

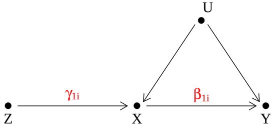
- medium sample size $\Rightarrow$ high variance
- small violations of assumptions $\Rightarrow$ large bias



Preview of Modern Approaches: Relaxing Constant Effects

Preview of Modern Approaches: Relaxing Constant Effects

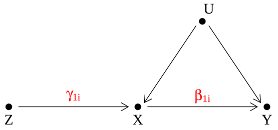
Suppose we believe that the effects of Z and X are different for different units.



$$Y_i = \beta_{0i} + \beta_{1i} X_i + U_i$$

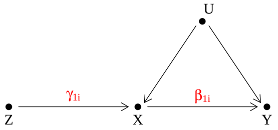
$$X_i = \gamma_{0i} + \gamma_{1i} Z_i + V_i$$

IV Estimator with Heterogeneous Effects



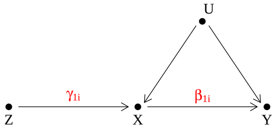
- regressing X on Z now only identifies $\overline{\gamma}_1$

IV Estimator with Heterogeneous Effects



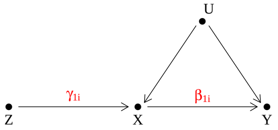
- regressing X on Z now only identifies $\overline{\gamma}_1$
- regressing Y on Z identifies only $\overline{\gamma_1 \cdot \beta_1}$

IV Estimator with Heterogeneous Effects



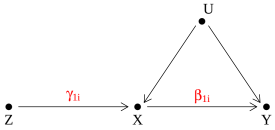
- regressing X on Z now only identifies $\overline{\gamma}_1$
- regressing Y on Z identifies only $\overline{\gamma_1 \cdot \beta_1}$
- $\overline{\gamma_1 \cdot \beta_1} \neq \overline{\gamma_1} \cdot \overline{\beta_1}$

IV Estimator with Heterogeneous Effects



- regressing X on Z now only identifies $\overline{\gamma}_1$
- regressing Y on Z identifies only $\overline{\gamma_1 \cdot \beta_1}$
- $\overline{\gamma_1 \cdot \beta_1} \neq \overline{\gamma_1} \cdot \overline{\beta_1}$
- Therefore the IV estimator does not estimate even the average β_1 ($\frac{\overline{\gamma_1 \cdot \beta_1}}{\overline{\gamma}_1} \neq \overline{\beta_1}$)

IV Estimator with Heterogeneous Effects



- regressing X on Z now only identifies $\overline{\gamma}_1$
- regressing Y on Z identifies only $\overline{\gamma_1 \cdot \beta_1}$
- $\overline{\gamma_1 \cdot \beta_1} \neq \overline{\gamma_1} \cdot \overline{\beta_1}$
- Therefore the IV estimator does not estimate even the average β_1 ($\frac{\overline{\gamma_1 \cdot \beta_1}}{\overline{\gamma}_1} \neq \overline{\beta_1}$)

With additional assumptions ($\gamma_{i1} \geq 0$ for all i), the IV estimator identifies a weighted average effect of X on Y according to the effects of Z on X.

Why IV?

Let's have a small notation change, and consider OLS:

Why IV?

Let's have a small notation change, and consider OLS:

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$

Why IV?

Let's have a small notation change, and consider OLS:

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the treatment variable (e.g. training)

Why IV?

Let's have a small notation change, and consider OLS:

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the treatment variable (e.g. training)
  - D may be endogenous so that $Cov[D, u_2] \neq 0$

Why IV?

Let's have a small notation change, and consider OLS:

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the treatment variable (e.g. training)
  - D may be endogenous so that $Cov[D, u_2] \neq 0$

- Recall that the OLS estimator for α_1 is given by:

$$\hat{\alpha}_{1,OLS} = \frac{\widehat{Cov}[Y,D]}{\widehat{Var}[D]} =$$

Why IV?

Let's have a small notation change, and consider OLS:

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the treatment variable (e.g. training)
  - D may be endogenous so that $Cov[D, u_2] \neq 0$

- Recall that the OLS estimator for α_1 is given by:

$$\hat{\alpha}_{1,OLS} = \frac{\widehat{Cov}[Y, D]}{\widehat{Var}[D]} = \frac{\widehat{Cov}[\alpha_0 + \alpha_1 D + u_2, D]}{\widehat{Var}[D]}$$

Why IV?

Let's have a small notation change, and consider OLS:

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the treatment variable (e.g. training)
  - D may be endogenous so that $Cov[D, u_2] \neq 0$

- Recall that the OLS estimator for α_1 is given by:

$$\hat{\alpha}_{1,OLS} = \frac{\widehat{Cov}[Y, D]}{\widehat{Var}[D]} = \frac{\widehat{Cov}[\alpha_0 + \alpha_1 D + u_2, D]}{\widehat{Var}[D]}$$

$$\hat{\alpha}_{1,OLS} = \frac{\alpha_1 \widehat{Var}[D] + \widehat{Cov}[D, u_2]}{\widehat{Var}[D]}$$

Why IV?

Let's have a small notation change, and consider OLS:

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the treatment variable (e.g. training)
  - D may be endogenous so that $Cov[D, u_2] \neq 0$
- Recall that the OLS estimator for α_1 is given by:

$$\hat{\alpha}_{1,OLS} = \frac{\widehat{Cov}[Y, D]}{\widehat{Var}[D]} = \frac{\widehat{Cov}[\alpha_0 + \alpha_1 D + u_2, D]}{\widehat{Var}[D]}$$

$$\hat{\alpha}_{1,OLS} = \frac{\alpha_1 \widehat{Var}[D] + \widehat{Cov}[D, u_2]}{\widehat{Var}[D]}$$

$$\hat{\alpha}_{1,OLS} = \alpha_1 + \frac{\widehat{Cov}[D, u_2]}{\widehat{Var}[D]}$$

Why IV?

Let's have a small notation change, and consider OLS:

- True model: $Y = \alpha_0 + \alpha_1 D + u_2$
  - D is the treatment variable (e.g. training)
  - D may be endogenous so that $Cov[D, u_2] \neq 0$

- Recall that the OLS estimator for α_1 is given by:

$$\begin{aligned}
\hat{\alpha}_{1,OLS} &= \frac{\widehat{Cov}[Y,D]}{\widehat{Var}[D]} = \frac{\widehat{Cov}[\alpha_0 + \alpha_1 D + u_2, D]}{\widehat{Var}[D]} \\
\hat{\alpha}_{1,OLS} &= \frac{\alpha_1 \widehat{Var}[D] + \widehat{Cov}[D, u_2]}{\widehat{Var}[D]} \\
\hat{\alpha}_{1,OLS} &= \alpha_1 + \frac{\widehat{Cov}[D, u_2]}{\widehat{Var}[D]} \\
E[\hat{\alpha}_{1,OLS}] &= \alpha_1 + E[\frac{\widehat{Cov}[D, u_2]}{\widehat{Var}[D]}]
\end{aligned}$$

so bias depends on correlation between u_2 and D

Instrumental Variable Estimator Assumptions

Instrumental Variable Estimator Assumptions

Imagine we have two equations:

Instrumental Variable Estimator Assumptions

Imagine we have two equations:

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$

Instrumental Variable Estimator Assumptions

Imagine we have two equations:

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
  - Z is our instrumental variable (e.g. randomized encouragement)
  - π_1 is effect of Z on D

Instrumental Variable Estimator Assumptions

Imagine we have two equations:

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
  - Z is our instrumental variable (e.g. randomized encouragement)
  - π_1 is effect of Z on D

Instrumental Variable Assumptions:

1. $\pi_1 \neq 0$ so Z creates some variation in D
   (called first stage or relevance)

Instrumental Variable Estimator Assumptions

Imagine we have two equations:

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
  - Z is our instrumental variable (e.g. randomized encouragement)
  - π_1 is effect of Z on D

Instrumental Variable Assumptions:

1. $\pi_1 \neq 0$ so Z creates some variation in D (called first stage or relevance)
2. Z is exogenous meaning $Cov[u_1, Z] = 0$ and $Cov[u_2, Z] = 0$.

Instrumental Variable Estimator Assumptions

Imagine we have two equations:

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
  - Z is our instrumental variable (e.g. randomized encouragement)
  - π_1 is effect of Z on D

Instrumental Variable Assumptions:

1. $\pi_1 \neq 0$ so Z creates some variation in D (called first stage or relevance)
2. Z is exogenous meaning $Cov[u_1, Z] = 0$ and $Cov[u_2, Z] = 0$. The latter is an exclusion restriction, it implies that the only reason why Z is correlated with Y is through the correlation between Z and D. So Z has no independent effect on Y.

Instrumental Variable Estimator Assumptions

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Instrumental Variable Estimator Assumptions

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Based on these IV assumptions we can identify three effects:

Instrumental Variable Estimator Assumptions

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Based on these IV assumptions we can identify three effects:

1. The first stage effect: Effect of Z on D.

Instrumental Variable Estimator Assumptions

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Based on these IV assumptions we can identify three effects:

1. The first stage effect: Effect of Z on D.
2. Reduced form or intent-to-treat effect: Effect of Z on Y.

Instrumental Variable Estimator Assumptions

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Based on these IV assumptions we can identify three effects:

1. The first stage effect: Effect of Z on D.
2. Reduced form or intent-to-treat effect: Effect of Z on Y.
3. The instrumental variable treatment effect: Effect of D on Y, using only the exogenous variation in D that is induced by Z.

First Stage Effect

First Stage Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

First Stage Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

First stage effect: Z on D

$$\hat{\pi}_1 = \frac{\widehat{Cov}[D, Z]}{\widehat{V}[Z]}$$

First Stage Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

First stage effect: Z on D

$$\hat{\pi}_1 = \frac{\widehat{Cov}[D, Z]}{\widehat{V}[Z]} = \frac{\widehat{Cov}[\pi_0 + \pi_1 Z + u_1, Z]}{\widehat{V}[Z]}$$

First Stage Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

First stage effect: Z on D

$$
\begin{aligned}
\hat{\pi}_1 &= \frac{\widehat{Cov}[D, Z]}{\widehat{V}[Z]} = \frac{\widehat{Cov}[\pi_0 + \pi_1 Z + u_1, Z]}{\widehat{V}[Z]} \\
\hat{\pi}_1 &= \frac{\pi_1 \widehat{Cov}[Z, Z] + \widehat{Cov}[Z, u_1]}{\widehat{V}[Z]}
\end{aligned}
$$

First Stage Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

First stage effect: Z on D

$$\hat{\pi}_1 = \frac{\widehat{Cov}[D, Z]}{\widehat{V}[Z]} = \frac{\widehat{Cov}[\pi_0 + \pi_1 Z + u_1, Z]}{\widehat{V}[Z]}$$

$$\hat{\pi}_1 = \frac{\pi_1 \widehat{Cov}[Z, Z] + \widehat{Cov}[Z, u_1]}{\widehat{V}[Z]}$$

$$\hat{\pi}_1 = \pi_1 + \frac{\widehat{Cov}[Z, u_1]}{\widehat{V}[Z]}$$

First Stage Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

First stage effect: Z on D

$$\hat{\pi}_1 = \frac{\widehat{Cov}[D, Z]}{\widehat{V}[Z]} = \frac{\widehat{Cov}[\pi_0 + \pi_1 Z + u_1, Z]}{\widehat{V}[Z]}$$

$$\hat{\pi}_1 = \frac{\pi_1 \widehat{Cov}[Z, Z] + \widehat{Cov}[Z, u_1]}{\widehat{V}[Z]}$$

$$\hat{\pi}_1 = \pi_1 + \frac{\widehat{Cov}[Z, u_1]}{\widehat{V}[Z]}$$

$$E[\hat{\pi}_1] = \pi_1 + E\left[\frac{\widehat{Cov}[Z, u_1]}{\widehat{V}[Z]}\right]$$

First Stage Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

First stage effect: Z on D

$$\begin{aligned}
\hat{\pi}_1 &= \frac{\widehat{Cov}[D, Z]}{\widehat{V}[Z]} = \frac{\widehat{Cov}[\pi_0 + \pi_1 Z + u_1, Z]}{\widehat{V}[Z]} \\
\hat{\pi}_1 &= \frac{\pi_1 \widehat{Cov}[Z, Z] + \widehat{Cov}[Z, u_1]}{\widehat{V}[Z]} \\
\hat{\pi}_1 &= \pi_1 + \frac{\widehat{Cov}[Z, u_1]}{\widehat{V}[Z]} \\
E[\hat{\pi}_1] &= \pi_1 + E\left[\frac{\widehat{Cov}[Z, u_1]}{\widehat{V}[Z]}\right] = \pi_1
\end{aligned}$$

$\hat{\pi}_1$ is consistent since $Cov[u_1, Z] = 0$

First Stage Effect in JTPA

First stage effect: Z on D: $\hat{\pi}_1 = \frac{\widehat{Cov}[D,Z]}{\widehat{V}[Z]}$

R Code

```
> cov(d[,c("earnings","training","assignmt")])
              earnings    training    assignmt
earnings 2.811338e+08 685.5254685 257.0625061
training 6.855255e+02   0.2456123   0.1390407
assignmt 2.570625e+02   0.1390407   0.221713
```

R Code

```
> 0.1390407/0.2217139
[1] 0.6271177
```

First Stage Effect in JTPA

R Code

```
> summary(lm(training~assignmt,data=d))

Call:
lm(formula = training ~ assignmt, data = d)

Residuals:
     Min       1Q   Median       3Q      Max
-0.64165 -0.01453 -0.01453  0.35835  0.98547

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.014528   0.006529   2.225   0.0261 *
assignmt    0.627118   0.007987  78.522   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.398 on 11202 degrees of freedom
Multiple R-squared: 0.355,        Adjusted R-squared: 0.355
F-statistic:  6166 on 1 and 11202 DF,  p-value: < 2.2e-1
```

Reduced Form/Intent-to-treat Effect

Reduced Form/Intent-to-treat Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Reduced Form/Intent-to-treat Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Reduced Form/Intent-to-treat Effect: Z on Y: Plug first into second stage:

$$Y = \alpha_0 + \alpha_1(\pi_0 + \pi_1 Z + u_1) + u_2$$

Reduced Form/Intent-to-treat Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Reduced Form/Intent-to-treat Effect: Z on Y: Plug first into second stage:

$$
\begin{aligned}
Y &= \alpha_0 + \alpha_1(\pi_0 + \pi_1 Z + u_1) + u_2 \\
Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2)
\end{aligned}
$$

Reduced Form/Intent-to-treat Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Reduced Form/Intent-to-treat Effect: Z on Y: Plug first into second stage:

$$
\begin{aligned}
Y &= \alpha_0 + \alpha_1(\pi_0 + \pi_1 Z + u_1) + u_2 \\
Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\
Y &= \gamma_0 + \gamma_1 Z + u_3
\end{aligned}
$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$.

Reduced Form/Intent-to-treat Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Reduced Form/Intent-to-treat Effect: Z on Y: Plug first into second stage:

$$\begin{aligned} Y &= \alpha_0 + \alpha_1(\pi_0 + \pi_1 Z + u_1) + u_2 \\ Y &= (\alpha_0 + \alpha_1\pi_0) + (\alpha_1\pi_1)Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1\pi_0$, $\gamma_1 = \alpha_1\pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Note that

$$\hat{\gamma}_1 = \frac{\widehat{Cov}[Y, Z]}{\widehat{Cov}[Z, Z]} = \frac{\widehat{Cov}[\gamma_0 + \gamma_1 Z + u_3, Z]}{\widehat{Cov}[Z, Z]}$$

Reduced Form/Intent-to-treat Effect

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Reduced Form/Intent-to-treat Effect: Z on Y: Plug first into second stage:

$$\begin{aligned} Y &= \alpha_0 + \alpha_1(\pi_0 + \pi_1 Z + u_1) + u_2 \\ Y &= (\alpha_0 + \alpha_1\pi_0) + (\alpha_1\pi_1)Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1\pi_0$, $\gamma_1 = \alpha_1\pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Note that

$$\begin{aligned} \hat{\gamma}_1 &= \frac{\widehat{Cov}[Y, Z]}{\widehat{Cov}[Z, Z]} = \frac{\widehat{Cov}[\gamma_0 + \gamma_1 Z + u_3, Z]}{\widehat{Cov}[Z, Z]} \\ E[\hat{\gamma}_1] &= \gamma_1 + E\left[\frac{\widehat{Cov}[Z, u_3]}{\widehat{Cov}[Z, Z]}\right] = \gamma_1 \end{aligned}$$

$\hat{\gamma}_1$ is consistent since $Cov[u_1, Z] = 0$ and $Cov[u_2, Z] = 0$ implies $Cov[u_3, Z] = 0$

Reduced Form/Intent-to-treat Effect

R Code

```
> summary(lm(earnings~assignmt,data=d))

Call:
lm(formula = earnings ~ assignmt, data = d)

Residuals:
   Min     1Q Median     3Q    Max
-16200 -13803  -4817   8950 139560

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15040.5      274.9  54.716  < 2e-16 ***
assignmt      1159.4      336.3   3.448 0.000567 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 16760 on 11202 degrees of freedom
Multiple R-squared: 0.00106,        Adjusted R-squared: 0.000971
F-statistic: 11.89 on 1 and 11202 DF,  p-value: 0.000566
```

Instrumental Variable Effect: Wald Estimator

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$Y = (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2)$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$
\begin{aligned}
Y &= (\alpha_0 + \alpha_1\pi_0) + (\alpha_1\pi_1)Z + (\alpha_1 u_1 + u_2) \\
Y &= \gamma_0 + \gamma_1 Z + u_3
\end{aligned}
$$

where $\gamma_0 = \alpha_0 + \alpha_1\pi_0$, $\gamma_1 = \alpha_1\pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$
\alpha_1 = \frac{\gamma_1}{\pi_1}
$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} =$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y, Z]/Cov[Z, Z]}{Cov[D, Z]/Cov[Z, Z]} = \frac{Cov[Y, Z]}{Cov[D, Z]}$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\begin{aligned} \alpha_1 &= \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y, Z]/Cov[Z, Z]}{Cov[D, Z]/Cov[Z, Z]} = \frac{Cov[Y, Z]}{Cov[D, Z]} \\ &= \frac{Cov[\alpha_0 + \alpha_1 D + u_2, Z]}{Cov[D, Z]} \end{aligned}$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\begin{aligned} \alpha_1 &= \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y, Z]/Cov[Z, Z]}{Cov[D, Z]/Cov[Z, Z]} = \frac{Cov[Y, Z]}{Cov[D, Z]} \\ &= \frac{Cov[\alpha_0 + \alpha_1 D + u_2, Z]}{Cov[D, Z]} = \frac{\alpha_1 Cov[D, Z] + Cov[u_2, Z]}{Cov[D, Z]} = \end{aligned}$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\begin{aligned} \alpha_1 &= \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y, Z]/Cov[Z, Z]}{Cov[D, Z]/Cov[Z, Z]} = \frac{Cov[Y, Z]}{Cov[D, Z]} \\ &= \frac{Cov[\alpha_0 + \alpha_1 D + u_2, Z]}{Cov[D, Z]} = \frac{\alpha_1 Cov[D, Z] + Cov[u_2, Z]}{Cov[D, Z]} = \alpha_1 + \frac{Cov[u_2, Z]}{Cov[D, Z]} \end{aligned}$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\begin{aligned} \alpha_1 &= \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y, Z]/Cov[Z, Z]}{Cov[D, Z]/Cov[Z, Z]} = \frac{Cov[Y, Z]}{Cov[D, Z]} \\ &= \frac{Cov[\alpha_0 + \alpha_1 D + u_2, Z]}{Cov[D, Z]} = \frac{\alpha_1 Cov[D, Z] + Cov[u_2, Z]}{Cov[D, Z]} = \alpha_1 + \frac{Cov[u_2, Z]}{Cov[D, Z]} \end{aligned}$$

$$E[\hat{\alpha}_1] =$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\begin{aligned} \alpha_1 &= \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y, Z]/Cov[Z, Z]}{Cov[D, Z]/Cov[Z, Z]} = \frac{Cov[Y, Z]}{Cov[D, Z]} \\ &= \frac{Cov[\alpha_0 + \alpha_1 D + u_2, Z]}{Cov[D, Z]} = \frac{\alpha_1 Cov[D, Z] + Cov[u_2, Z]}{Cov[D, Z]} = \alpha_1 + \frac{Cov[u_2, Z]}{Cov[D, Z]} \\ E[\hat{\alpha}_1] &= \alpha_1 + E\left[\frac{\widehat{Cov}[u_2, Z]}{\widehat{Cov}[D, Z]}\right] \end{aligned}$$

Instrumental Variable Effect: Wald Estimator

- Second Stage: $Y = \alpha_0 + \alpha_1 D + u_2$
- First Stage: $D = \pi_0 + \pi_1 Z + u_1$
- IV assumptions: $Cov[u_1, Z] = 0$, $\pi_1 \neq 0$, and $Cov[u_2, Z] = 0$

IV Effect: D on Y using exogenous variation in D that is induced by Z. Recall

$$\begin{aligned} Y &= (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \\ Y &= \gamma_0 + \gamma_1 Z + u_3 \end{aligned}$$

where $\gamma_0 = \alpha_0 + \alpha_1 \pi_0$, $\gamma_1 = \alpha_1 \pi_1$, and $u_3 = \alpha_1 u_1 + u_2$. Given this, we can identify α_1:

$$\begin{aligned} \alpha_1 &= \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y,Z]/Cov[Z,Z]}{Cov[D,Z]/Cov[Z,Z]} = \frac{Cov[Y,Z]}{Cov[D,Z]} \\ &= \frac{Cov[\alpha_0 + \alpha_1 D + u_2, Z]}{Cov[D,Z]} = \frac{\alpha_1 Cov[D,Z] + Cov[u_2,Z]}{Cov[D,Z]} = \alpha_1 + \frac{Cov[u_2,Z]}{Cov[D,Z]} \\ E[\hat{\alpha}_1] &= \alpha_1 + E\left[\frac{\widehat{Cov}[u_2,Z]}{\widehat{Cov}[D,Z]}\right] \end{aligned}$$

$\hat{\alpha}_1$ is consistent if $Cov[u_2, Z] = 0$ but has a bias which decreases with instrument strength.

Instrumental Variable Effect: Wald Estimator

Instrumental Variable Effect: $\alpha_1 = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on D}} = \frac{Cov[Y,Z]}{Cov[D,Z]}$

R Code

```
> cov(d[,c("earnings","training","assignmt")])
             earnings    training    assignmt
earnings 2.811338e+08 685.5254685 257.0625061
training 6.855255e+02   0.2456123   0.1390407
assignmt 2.570625e+02   0.1390407   0.221713
```

R Code

```
> 257.0625061/0.1390407
[1] 1848.829
```

Instrumental Variable Effect: Two Stage Least Squares

Instrumental Variable Effect: Two Stage Least Squares

The instrumental variable estimator:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{Cov[Y, Z]}{Cov[D, Z]}$$

is numerically equivalent to the following two step procedure:

Instrumental Variable Effect: Two Stage Least Squares

The instrumental variable estimator:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{Cov[Y, Z]}{Cov[D, Z]}$$

is numerically equivalent to the following two step procedure:

1. Fit first stage and obtain fitted values $\hat{D} = \hat{\pi}_0 + \hat{\pi}_1 Z$

Instrumental Variable Effect: Two Stage Least Squares

The instrumental variable estimator:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{Cov[Y, Z]}{Cov[D, Z]}$$

is numerically equivalent to the following two step procedure:

1. Fit first stage and obtain fitted values $\hat{D} = \hat{\pi}_0 + \hat{\pi}_1 Z$
2. Plug into second stage:

Instrumental Variable Effect: Two Stage Least Squares

The instrumental variable estimator:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{Cov[Y, Z]}{Cov[D, Z]}$$

is numerically equivalent to the following two step procedure:

1. Fit first stage and obtain fitted values $\hat{D} = \hat{\pi}_0 + \hat{\pi}_1 Z$
2. Plug into second stage:

$$Y = \alpha_0 + \alpha_1 \hat{D} + u_2$$

Instrumental Variable Effect: Two Stage Least Squares

The instrumental variable estimator:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{Cov[Y,Z]}{Cov[D,Z]}$$

is numerically equivalent to the following two step procedure:

1. Fit first stage and obtain fitted values $\hat{D} = \hat{\pi}_0 + \hat{\pi}_1 Z$
2. Plug into second stage:

$$\begin{aligned} Y &= \alpha_0 + \alpha_1 \hat{D} + u_2 \\ Y &= \alpha_0 + \alpha_1 (\hat{\pi}_0 + \hat{\pi}_1 Z) + u_2 \end{aligned}$$

Instrumental Variable Effect: Two Stage Least Squares

The instrumental variable estimator:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{Cov[Y, Z]}{Cov[D, Z]}$$

is numerically equivalent to the following two step procedure:

1. Fit first stage and obtain fitted values $\hat{D} = \hat{\pi}_0 + \hat{\pi}_1 Z$
2. Plug into second stage:

$$\begin{aligned}
Y &= \alpha_0 + \alpha_1 \hat{D} + u_2 \\
Y &= \alpha_0 + \alpha_1(\hat{\pi}_0 + \hat{\pi}_1 Z) + u_2 \\
Y &= (\alpha_0 + \alpha_1 \hat{\pi}_0) + \alpha_1(\hat{\pi}_1 Z) + u_2
\end{aligned}$$

Instrumental Variable Effect: Two Stage Least Squares

The instrumental variable estimator:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{Cov[Y, Z]}{Cov[D, Z]}$$

is numerically equivalent to the following two step procedure:

1. Fit first stage and obtain fitted values $\hat{D} = \hat{\pi}_0 + \hat{\pi}_1 Z$
2. Plug into second stage:

$$\begin{aligned} Y &= \alpha_0 + \alpha_1 \hat{D} + u_2 \\ Y &= \alpha_0 + \alpha_1(\hat{\pi}_0 + \hat{\pi}_1 Z) + u_2 \\ Y &= (\alpha_0 + \alpha_1 \hat{\pi}_0) + \alpha_1(\hat{\pi}_1 Z) + u_2 \end{aligned}$$

- α_1 is solely identified based on variation in D that comes from Z

Instrumental Variable Effect: Two Stage Least Squares

The instrumental variable estimator:

$$\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{Cov[Y, Z]}{Cov[D, Z]}$$

is numerically equivalent to the following two step procedure:

1. Fit first stage and obtain fitted values $\hat{D} = \hat{\pi}_0 + \hat{\pi}_1 Z$
2. Plug into second stage:

$$\begin{aligned} Y &= \alpha_0 + \alpha_1 \hat{D} + u_2 \\ Y &= \alpha_0 + \alpha_1 (\hat{\pi}_0 + \hat{\pi}_1 Z) + u_2 \\ Y &= (\alpha_0 + \alpha_1 \hat{\pi}_0) + \alpha_1 (\hat{\pi}_1 Z) + u_2 \end{aligned}$$

- α_1 is solely identified based on variation in D that comes from Z
- Point estimates from second regression are equivalent to IV estimator, the standard errors are not quite correct since they ignore the estimation uncertainty in $\hat{\pi}_0$ and $\hat{\pi}_1$.

Instrumental Variable Effect: Two Stage Least Squares

R Code

```
> training_hat <- lm(training~assignmt,data=d)$fitted
> summary(lm(earnings~training_hat,data=d))

Call:
lm(formula = earnings ~ training_hat, data = d)

Residuals:
   Min     1Q Median     3Q    Max
-16200 -13803  -4817   8950 139560

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   15013.6      281.3  53.375  < 2e-16 ***
training_hat   1848.8      536.2   3.448 0.000567 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: 16760 on 11202 degrees of freedom
Multiple R-squared: 0.00106,       Adjusted R-squared: 0.000971
F-statistic: 11.89 on 1 and 11202 DF,  p-value: 0.0005669
```

Instrumental Variable Effect: Two Stage Least Squares

R Code

```
> library(AER)
> summary(ivreg(earnings ~ training | assignmt,data = d))
Call:
ivreg(formula = earnings ~ training | assignmt, data = d)
Residuals:
   Min     1Q Median     3Q    Max
-16862 -13716  -4943   8834 140746
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15013.6      280.6  53.508  < 2e-16 ***
training      1848.8      534.9   3.457 0.000549 ***
---
Residual standard error: 16720 on 11202 degrees of freedom
Multiple R-Squared: 0.00603,        Adjusted R-squared: 0.005941
Wald test: 11.95 on 1 and 11202 DF,  p-value: 0.0005491
```

Judging the Credibility of IV Estimates

Judging the Credibility of IV Estimates

- The probability limit of the IV estimator is given by:

$$plim\, \alpha_{D,IV} = \alpha_D + \frac{Corr(Z, u_2)}{Corr(Z, D)} \frac{\sigma^{u_2}}{\sigma^D}$$

  so to obtain consistent estimates the instrument Z must:

Judging the Credibility of IV Estimates

- The probability limit of the IV estimator is given by:

$$plim\, \alpha_{D,IV} = \alpha_D + \frac{Corr(Z, u_2)}{Corr(Z, D)} \frac{\sigma^{u_2}}{\sigma^D}$$

  so to obtain consistent estimates the instrument Z must:

- Be Relevant: $Cov(Z, D) \neq 0$ (testable)

Judging the Credibility of IV Estimates

- The probability limit of the IV estimator is given by:

$$plim\, \alpha_{D,IV} = \alpha_D + \frac{Corr(Z, u_2)}{Corr(Z, D)} \frac{\sigma^{u_2}}{\sigma^D}$$

  so to obtain consistent estimates the instrument Z must:

- Be Relevant: $Cov(Z, D) \neq 0$ (testable)
  - if $Cov(Z, D)$ is small, the instrument is weak.

Judging the Credibility of IV Estimates

- The probability limit of the IV estimator is given by:

$$plim\, \alpha_{D,IV} = \alpha_D + \frac{Corr(Z, u_2)}{Corr(Z, D)} \frac{\sigma^{u_2}}{\sigma^D}$$

  so to obtain consistent estimates the instrument Z must:

- Be Relevant: $Cov(Z, D) \neq 0$ (testable)
  - if $Cov(Z, D)$ is small, the instrument is weak.
  - weak instruments increases bias, but estimator remains consistent.

Judging the Credibility of IV Estimates

- The probability limit of the IV estimator is given by:

$$plim\, \alpha_{D,IV} = \alpha_D + \frac{Corr(Z, u_2)}{Corr(Z, D)} \frac{\sigma^{u_2}}{\sigma^D}$$

  so to obtain consistent estimates the instrument Z must:

- Be Relevant: $Cov(Z, D) \neq 0$ (testable)
  - if $Cov(Z, D)$ is small, the instrument is weak.
  - weak instruments increases bias, but estimator remains consistent.
  - bias can be substantial even for very large sample sizes when the instrument is weak.

Judging the Credibility of IV Estimates

- The probability limit of the IV estimator is given by:

$$plim\, \alpha_{D,IV} = \alpha_D + \frac{Corr(Z, u_2)}{Corr(Z, D)} \frac{\sigma^{u_2}}{\sigma^D}$$

  so to obtain consistent estimates the instrument Z must:

- Be Relevant: $Cov(Z, D) \neq 0$ (testable)
  - if $Cov(Z, D)$ is small, the instrument is weak.
  - weak instruments increases bias, but estimator remains consistent.
  - bias can be substantial even for very large sample sizes when the instrument is weak.
- Satisfy Exclusion Restriction: $Cov(Z, u_2) = 0$ (untestable)

Judging the Credibility of IV Estimates

- The probability limit of the IV estimator is given by:

$$plim\, \alpha_{D,IV} = \alpha_D + \frac{Corr(Z, u_2)}{Corr(Z, D)} \frac{\sigma^{u_2}}{\sigma^D}$$

  so to obtain consistent estimates the instrument Z must:

- Be Relevant: $Cov(Z, D) \neq 0$ (testable)
  - if $Cov(Z, D)$ is small, the instrument is weak.
  - weak instruments increases bias, but estimator remains consistent.
  - bias can be substantial even for very large sample sizes when the instrument is weak.
- Satisfy Exclusion Restriction: $Cov(Z, u_2) = 0$ (untestable)
  - if Z has an independent effect on Y other than through D we have $Cov(Z, u_2) \neq 0$

Judging the Credibility of IV Estimates

- The probability limit of the IV estimator is given by:

$$plim\, \alpha_{D,IV} = \alpha_D + \frac{Corr(Z, u_2)}{Corr(Z, D)} \frac{\sigma^{u_2}}{\sigma^D}$$

  so to obtain consistent estimates the instrument Z must:

- Be Relevant: $Cov(Z, D) \neq 0$ (testable)
  - if $Cov(Z, D)$ is small, the instrument is weak.
  - weak instruments increases bias, but estimator remains consistent.
  - bias can be substantial even for very large sample sizes when the instrument is weak.
- Satisfy Exclusion Restriction: $Cov(Z, u_2) = 0$ (untestable)
  - if Z has an independent effect on Y other than through D we have $Cov(Z, u_2) \neq 0$
  - if assumption not met estimates are inconsistent

Judging the Credibility of IV Estimates

- The probability limit of the IV estimator is given by:

$$plim\, \alpha_{D,IV} = \alpha_D + \frac{Corr(Z, u_2)}{Corr(Z, D)} \frac{\sigma^{u_2}}{\sigma^D}$$

  so to obtain consistent estimates the instrument Z must:

- Be Relevant: $Cov(Z, D) \neq 0$ (testable)
  - if $Cov(Z, D)$ is small, the instrument is weak.
  - weak instruments increases bias, but estimator remains consistent.
  - bias can be substantial even for very large sample sizes when the instrument is weak.
- Satisfy Exclusion Restriction: $Cov(Z, u_2) = 0$ (untestable)
  - if Z has an independent effect on Y other than through D we have $Cov(Z, u_2) \neq 0$
  - if assumption not met estimates are inconsistent
  - small violations can lead to significant large sample bias unless instruments are strong

Judging the Credibility of IV Estimates

- The probability limit of the IV estimator is given by:

$$plim\, \alpha_{D,IV} = \alpha_D + \frac{Corr(Z, u_2)}{Corr(Z, D)} \frac{\sigma^{u_2}}{\sigma^D}$$

  so to obtain consistent estimates the instrument Z must:

- Be Relevant: $Cov(Z, D) \neq 0$ (testable)
  - if $Cov(Z, D)$ is small, the instrument is weak.
  - weak instruments increases bias, but estimator remains consistent.
  - bias can be substantial even for very large sample sizes when the instrument is weak.
- Satisfy Exclusion Restriction: $Cov(Z, u_2) = 0$ (untestable)
  - if Z has an independent effect on Y other than through D we have $Cov(Z, u_2) \neq 0$
  - if assumption not met estimates are inconsistent
  - small violations can lead to significant large sample bias unless instruments are strong
- Failure of either condition is a problem! But both conditions can be hard to satisfy at the same time. There often is a tradeoff.

When Should We Believe The Exclusion Restriction?

- Does a randomly assigned instrument Z always satisfy $Cov(Z, u_2) = 0$?

When Should We Believe The Exclusion Restriction?

- Does a randomly assigned instrument Z always satisfy $Cov(Z, u_2) = 0$?
- No!

When Should We Believe The Exclusion Restriction?

- Does a randomly assigned instrument Z always satisfy $Cov(Z, u_2) = 0$?
- No! Encouragement may still have independent effect on outcome other than through the treatment.

When Should We Believe The Exclusion Restriction?

- Does a randomly assigned instrument Z always satisfy $Cov(Z, u_2) = 0$?
- No! Encouragement may still have independent effect on outcome other than through the treatment.For example,
  - being draft eligible might encourage people to go to school and that might impact their earnings (Angrist 1990, 330)

When Should We Believe The Exclusion Restriction?

- Does a randomly assigned instrument Z always satisfy $Cov(Z, u_2) = 0$?
- No! Encouragement may still have independent effect on outcome other than through the treatment.For example,
  - being draft eligible might encourage people to go to school and that might impact their earnings (Angrist 1990, 330)
  - having a low draft number might encourage an employer to invest less in the employee because of concerns they will be drafted (Wooldridge 2010, 94)

When Should We Believe The Exclusion Restriction?

- Does a randomly assigned instrument Z always satisfy $Cov(Z, u_2) = 0$?
- No! Encouragement may still have independent effect on outcome other than through the treatment.For example,
  - being draft eligible might encourage people to go to school and that might impact their earnings (Angrist 1990, 330)
  - having a low draft number might encourage an employer to invest less in the employee because of concerns they will be drafted (Wooldridge 2010, 94)
- When designing an encouragement experiment we need to be careful to design encouragements so that they are relevant, but also narrowly constructed to only create variation in treatment intake.

When Should We Believe The Exclusion Restriction?

- Does a randomly assigned instrument Z always satisfy $Cov(Z, u_2) = 0$?
- No! Encouragement may still have independent effect on outcome other than through the treatment.For example,
  - being draft eligible might encourage people to go to school and that might impact their earnings (Angrist 1990, 330)
  - having a low draft number might encourage an employer to invest less in the employee because of concerns they will be drafted (Wooldridge 2010, 94)
- When designing an encouragement experiment we need to be careful to design encouragements so that they are relevant, but also narrowly constructed to only create variation in treatment intake.
- In observational work, imagining the ideal experiment (and associated compliance problem) can be helpful.

When Should We Believe The Exclusion Restriction?

- Does a randomly assigned instrument Z always satisfy $Cov(Z, u_2) = 0$?
- No! Encouragement may still have independent effect on outcome other than through the treatment.For example,
  - being draft eligible might encourage people to go to school and that might impact their earnings (Angrist 1990, 330)
  - having a low draft number might encourage an employer to invest less in the employee because of concerns they will be drafted (Wooldridge 2010, 94)
- When designing an encouragement experiment we need to be careful to design encouragements so that they are relevant, but also narrowly constructed to only create variation in treatment intake.
- In observational work, imagining the ideal experiment (and associated compliance problem) can be helpful.
- Requires understanding of the context!

General Words of Caution

General Words of Caution

These methods are not a panacea. Even if someone calls them a “natural experiment” that doesn’t actually make it like an experiment.

General Words of Caution

These methods are not a panacea. Even if someone calls them a "natural experiment" that doesn't actually make it like an experiment.

> *"The general lesson is once again the ultimate futility of trying to avoid thinking about how and why things work"*
> *- Angus Deaton (2010)*

General Words of Caution

These methods are not a panacea. Even if someone calls them a "natural experiment" that doesn't actually make it like an experiment.

> *"The general lesson is once again the ultimate futility of trying to avoid thinking about how and why things work"*
> *- Angus Deaton (2010)*

> *"[there is a] risk [of] transforming the methodologic dream of avoiding unmeasured confounding into a nightmare of conflicting biased estimates"*
> *- Hernán and Robins (2006)*

Conclusion

Conclusion

- Core logic of IV: inflate the intention to treat effect by the inverse of the compliance.

Conclusion

- Core logic of IV: inflate the intention to treat effect by the inverse of the compliance.
- IV works only under very specific circumstances
(e.g. well designed encouragement design experiments)

Conclusion

- Core logic of IV: inflate the intention to treat effect by the inverse of the compliance.
- IV works only under very specific circumstances
  (e.g. well designed encouragement design experiments)
- Often, it will be difficult to find instruments that are both relevant (strong enough) and satisfy the exclusion restriction

Conclusion

- Core logic of IV: inflate the intention to treat effect by the inverse of the compliance.
- IV works only under very specific circumstances
  (e.g. well designed encouragement design experiments)
- Often, it will be difficult to find instruments that are both relevant (strong enough) and satisfy the exclusion restriction
- Violations of assumptions can lead to large biases and estimation theory is complicated

Conclusion

- Core logic of IV: inflate the intention to treat effect by the inverse of the compliance.
- IV works only under very specific circumstances (e.g. well designed encouragement design experiments)
- Often, it will be difficult to find instruments that are both relevant (strong enough) and satisfy the exclusion restriction
- Violations of assumptions can lead to large biases and estimation theory is complicated
- So far, we have assumed constant treatment effects α_D which seems unrealistic in most settings. Often an instrument affects only a subpopulation of interest and we have little information about treatment effects for other units that may not be affected by the instrument at all.

Conclusion

- Core logic of IV: inflate the intention to treat effect by the inverse of the compliance.
- IV works only under very specific circumstances
(e.g. well designed encouragement design experiments)
- Often, it will be difficult to find instruments that are both relevant (strong enough) and satisfy the exclusion restriction
- Violations of assumptions can lead to large biases and estimation theory is complicated
- So far, we have assumed constant treatment effects α_D which seems unrealistic in most settings. Often an instrument affects only a subpopulation of interest and we have little information about treatment effects for other units that may not be affected by the instrument at all.
- Next time we'll discuss modern IV with heterogeneous potential outcomes

Fun With Coarsening Bias

Coarsening Bias: How Coarse Treatment Measurement Upwardly Biases Instrumental Variable Estimates

John Marshall
Department of Government, Harvard University, Cambridge, MA 02138
e-mail: jlmarsh@fas.harvard.edu (corresponding author)

Edited by Jonathan Katz

Political scientists increasingly use instrumental variable (IV) methods, and must often choose between operationalizing their endogenous treatment variable as discrete or continuous. For theoretical and data availability reasons, researchers frequently coarsen treatments with multiple intensities (e.g., treating a continuous treatment as binary). I show how such coarsening can substantially upwardly bias IV estimates by subtly violating the exclusion restriction assumption, and demonstrate that the extent of this bias depends upon the first stage and underlying causal response function. However, standard IV methods using a treatment where multiple intensities are affected by the instrument—even when fine-grained measurement at every intensity is not possible—recover a consistent causal estimate without requiring a stronger exclusion restriction assumption. These analytical insights are illustrated in the context of identifying the long-run effect of high school education on voting Conservative in Great Britain. I demonstrate that coarsening years of schooling into an indicator for completing high school upwardly biases the IV estimate by a factor of three.

The Idea

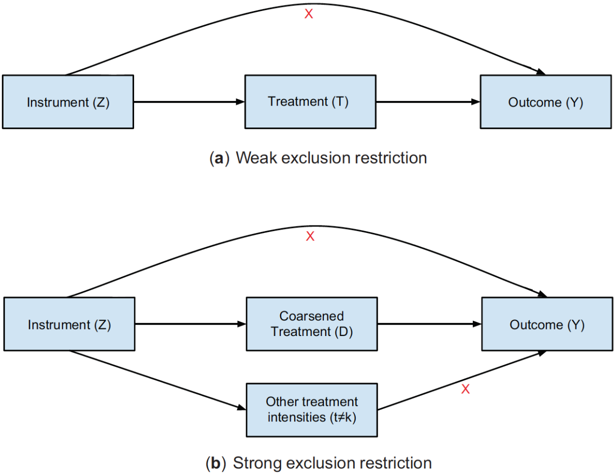


Fig. 1 Graphical representation of weak and strong exclusion restrictions.

Design

- Data: British Election Survey 1979-2010
- Outcome: voting for conservative party in most recent election
- Instrument: respondents turning 14 in 1947 or later who were affected by the 1947 school leaving reform (increased age from 14 to 15)
- Treatment: either years of schooling or coarsened indicator for completed high school or not

Data

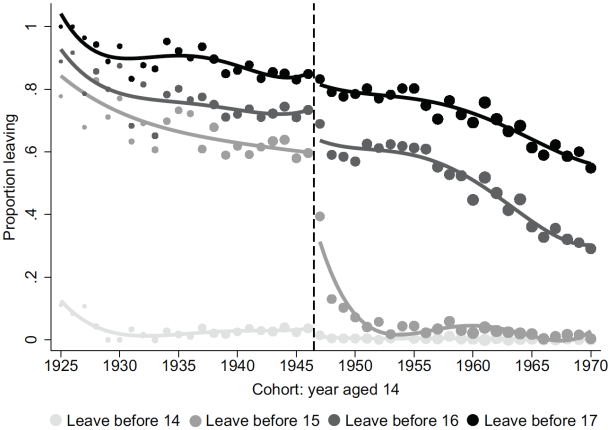


Fig. 3 1947 compulsory schooling reform and student leaving age by cohort.
Notes: Data are from the British Election Survey. Curves represent fourth-order polynomial fits. Gray dots are birth-year cohort averages, and their size reflects their weight in the sample.

Findings

- Finding: Using the dichotomous version of the treatment inflates the result by a factor of three
- Suggestion: Use the linear version of the treatment (although see the article for more details!)

Where We've Been and Where We're Going...

Where We've Been and Where We're Going...

- Last Week
  - selection on observables and measured confounding
- This Week
  - Monday:
    - natural experiments
    - classical view of instrumental variables
  - Wednesday:
    - modern view of instrumental variables
    - regression discontinuity
- The Following Week
  - repeated observations
- Long Run
  - causality with measured confounding $\rightarrow$ unmeasured confounding $\rightarrow$ repeated data

Questions?

Identification with Traditional Instrumental Variables

- Two equations:
  - $Y = \gamma + \alpha D + \varepsilon$ (Second Stage)
  - $D = \tau + \rho Z + \eta$ (First Stage)
- Four Assumptions
  1. Exogeneity: $Cov(Z, \eta) = 0$
  2. Exclusion: $Cov(Z, \varepsilon) = 0$
  3. First Stage Relevance: $\rho \neq 0$
  4. Homogeneity: $\alpha = Y_{1,i} - Y_{0,i}$ constant for all units i.
     Or in the case of a multivalued treatment with s levels we assume $\alpha = Y_{s,i} - Y_{s-1,i}$.

Instrumental Variables and Potential Outcomes

- Basic idea of IV:
  - D_i not randomized, but Z_i is
  - Z_i only affects Y_i through D_i

Instrumental Variables and Potential Outcomes

- Basic idea of IV:
  - D_i not randomized, but Z_i is
  - Z_i only affects Y_i through D_i
- D_i now depends on $Z_i \rightsquigarrow$ two potential treatments: $D_i(1) = D_i(z = 1)$ and $D_i(0)$.

Instrumental Variables and Potential Outcomes

- Basic idea of IV:
  - D_i not randomized, but Z_i is
  - Z_i only affects Y_i through D_i
- D_i now depends on $Z_i \rightsquigarrow$ two potential treatments: $D_i(1) = D_i(z = 1)$ and $D_i(0)$.
- Outcome can depend on both the treatment and the instrument: $Y_i(d, z)$ is the outcome if unit i had received treatment $D_i = d$ and instrument value $Z_i = z$.

Potential Outcome Model for Instrumental Variables

Potential Outcome Model for Instrumental Variables

Definition (Instrument)

Z_i: Binary instrument for unit i.

$$Z_i = \begin{cases} 1 & \text{if unit } i \text{ "encouraged" to receive treatment} \\ 0 & \text{if unit } i \text{ "encouraged" to receive control} \end{cases}$$

Definition (Potential Treatments)

$D(z)$ indicates potential treatment status given $Z = z$

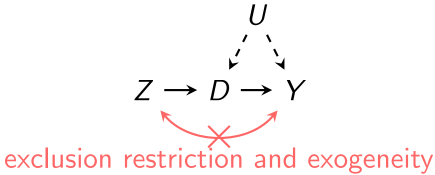
- $D_i(1) = 1$ encouraged to take treatment and takes treatment

Assumption

Observed treatments are realized as

$$D_i = Z_i \cdot D_i(1) + (1 - Z_i) \cdot D_i(0) \text{ so } D_i = \begin{cases} D_i(1) & \text{if } Z_i = 1 \\ D_i(0) & \text{if } Z_i = 0 \end{cases}$$

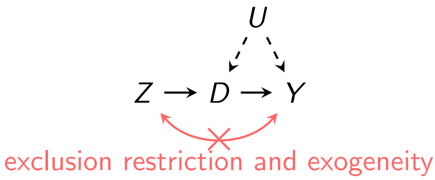
Key Assumptions in the Modern Approach



Assumptions:

1. Exogeneity of the Instrument

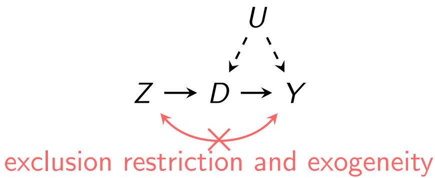
Key Assumptions in the Modern Approach



Assumptions:

1. Exogeneity of the Instrument
2. Exclusion Restriction

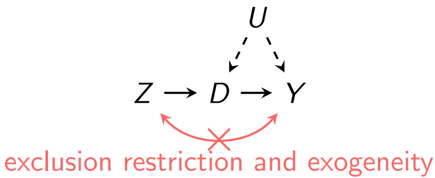
Key Assumptions in the Modern Approach



Assumptions:

1. Exogeneity of the Instrument
2. Exclusion Restriction
3. First-stage relationship

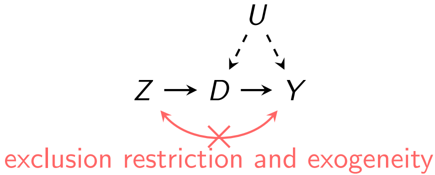
Key Assumptions in the Modern Approach



Assumptions:

1. Exogeneity of the Instrument
2. Exclusion Restriction
3. First-stage relationship
4. Monotonicity

Key Assumptions in the Modern Approach



Assumptions:

1. Exogeneity of the Instrument
2. Exclusion Restriction
3. First-stage relationship
4. Monotonicity

You may sometimes see assumptions 1 and 2 collapsed into an assumption called something like "Ignorability of the Instrument". I find it helpful to assess them separately though.

Assumption 1: Exogeneity of the Instrument

- Essentially we want the instrument to be randomized:

$$[\{Y_i(d,z), \forall d,z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$$

Assumption 1: Exogeneity of the Instrument

- Essentially we want the instrument to be randomized:

$$[\{Y_i(d,z), \forall d,z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$$

- We can weaken this to conditional ignorability. But why believe conditional ignorability for the instrument but not the treatment?

Assumption 1: Exogeneity of the Instrument

- Essentially we want the instrument to be randomized:

$$[\{Y_i(d,z), \forall d,z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$$

- We can weaken this to conditional ignorability. But why believe conditional ignorability for the instrument but not the treatment?
- Best instruments are truly randomized.

Assumption 1: Exogeneity of the Instrument

- Essentially we want the instrument to be randomized:

$$[\{Y_i(d,z), \forall d, z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$$

- We can weaken this to conditional ignorability. But why believe conditional ignorability for the instrument but not the treatment?
- Best instruments are truly randomized.
- This assumption alone gets us the intent-to-treat (ITT) effect:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)]$$

Assumption 1: Exogeneity of the Instrument

- Essentially we want the instrument to be randomized:

$$[\{Y_i(d,z), \forall d, z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$$

- We can weaken this to conditional ignorability. But why believe conditional ignorability for the instrument but not the treatment?
- Best instruments are truly randomized.
- This assumption alone gets us the intent-to-treat (ITT) effect:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_i(D_i(1), 1) - Y_i(D_i(0), 0)]$$

- Sometimes the ITT is interesting in its own right and should probably be reported regardless.

Assumption 2: Exclusion Restriction

- The instrument has no direct effect on the outcome, once we fix the value of the treatment.

$$Y_i(d, 1) = Y_i(d, 0) \quad \text{for } d = 0, 1$$

Assumption 2: Exclusion Restriction

- The instrument has no direct effect on the outcome, once we fix the value of the treatment.

$$Y_i(d, 1) = Y_i(d, 0) \quad \text{for } d = 0, 1$$

- Given this exclusion restriction, we know that the potential outcomes for each treatment status only depend on the treatment, not the instrument:

$$Y_i(1) \equiv Y_i(1, 1) = Y_i(1, 0)$$
$$Y_i(0) \equiv Y_i(0, 1) = Y_i(0, 0)$$

Assumption 2: Exclusion Restriction

- The instrument has no direct effect on the outcome, once we fix the value of the treatment.

$$Y_i(d, 1) = Y_i(d, 0) \quad \text{for } d = 0, 1$$

- Given this exclusion restriction, we know that the potential outcomes for each treatment status only depend on the treatment, not the instrument:

$$Y_i(1) \equiv Y_i(1, 1) = Y_i(1, 0)$$
$$Y_i(0) \equiv Y_i(0, 1) = Y_i(0, 0)$$

- Random assignment of the instrument is not sufficient for exclusion

Assumption 2: Exclusion Restriction

- The instrument has no direct effect on the outcome, once we fix the value of the treatment.

$$Y_i(d,1) = Y_i(d,0) \quad \text{for } d = 0,1$$

- Given this exclusion restriction, we know that the potential outcomes for each treatment status only depend on the treatment, not the instrument:

$$Y_i(1) \equiv Y_i(1,1) = Y_i(1,0)$$
$$Y_i(0) \equiv Y_i(0,1) = Y_i(0,0)$$

- Random assignment of the instrument is not sufficient for exclusion
- NOT A TESTABLE ASSUMPTION

Assumption 3: First Stage Relationship

- The instrument must have an effect on the treatment.

$$E[D_i(1) - D_i(0)] \neq 0$$

Assumption 3: First Stage Relationship

- The instrument must have an effect on the treatment.

$$E[D_i(1) - D_i(0)] \neq 0$$

- Implies that
  - $\text{Cov}(D, Z) \neq 0$ (instrument and treatment are linearly related)
  - $0 < P(Z_i = 1) < 1$ (all units have some chance of getting instrument)
  - $P(D(1) = 1) \neq P(D(0) = 1)$ (proportion treated would be different if all received instrument than if all didn't)

Assumption 3: First Stage Relationship

- The instrument must have an effect on the treatment.

$$E[D_i(1) - D_i(0)] \neq 0$$

- Implies that
  - $\text{Cov}(D, Z) \neq 0$ (instrument and treatment are linearly related)
  - $0 < P(Z_i = 1) < 1$ (all units have some chance of getting instrument)
  - $P(D(1) = 1) \neq P(D(0) = 1)$ (proportion treated would be different if all received instrument than if all didn't)
- This is testable by regressing D on Z (or making a scatter plot of D and Z)

Assumption 3: First Stage Relationship

- The instrument must have an effect on the treatment.

$$E[D_i(1) - D_i(0)] \neq 0$$

- Implies that
  - $\text{Cov}(D, Z) \neq 0$ (instrument and treatment are linearly related)
  - $0 < P(Z_i = 1) < 1$ (all units have some chance of getting instrument)
  - $P(D(1) = 1) \neq P(D(0) = 1)$ (proportion treated would be different if all received instrument than if all didn't)
- This is testable by regressing D on Z (or making a scatter plot of D and Z)
- Note that the finite-sample bias of the IV estimator depends inversely on the strength of the instrument. Thus, for practical sample sizes you need a strong first stage effect.

Assumption 4: Monotonicity

- To allow for heterogenous effects we need to make a new assumption about the relationship between the instrument and the treatment.

Assumption 4: Monotonicity

- To allow for heterogenous effects we need to make a new assumption about the relationship between the instrument and the treatment.
- Monotonicity says that the presence of the instrument never dissuades someone from taking the treatment:

$$D_i(1) - D_i(0) \geq 0$$

Assumption 4: Monotonicity

- To allow for heterogenous effects we need to make a new assumption about the relationship between the instrument and the treatment.
- Monotonicity says that the presence of the instrument never dissuades someone from taking the treatment:

$$D_i(1) - D_i(0) \geq 0$$

- Note if this holds in the opposite direction $D_i(1) - D_i(0) \leq 0$, we can always rescale D_i to make the assumption hold.

Principal Strata

Principal Strata

Following Angrist, Imbens, and Rubin (1996), we can define four subpopulations (for cases with a binary treatment and a binary instrument):

Principal Strata

Following Angrist, Imbens, and Rubin (1996), we can define four subpopulations (for cases with a binary treatment and a binary instrument):

Definition

- Compliers: $D_i(1) > D_i(0)$ ($D_i(0) = 0$ and $D_i(1) = 1$).

Principal Strata

Following Angrist, Imbens, and Rubin (1996), we can define four subpopulations (for cases with a binary treatment and a binary instrument):

Definition

- Compliers: $D_i(1) > D_i(0)$ ($D_i(0) = 0$ and $D_i(1) = 1$).
- Always-takers: $D_i(1) = D_i(0) = 1$.

Principal Strata

Following Angrist, Imbens, and Rubin (1996), we can define four subpopulations (for cases with a binary treatment and a binary instrument):

Definition

- Compliers: $D_i(1) > D_i(0)$ ($D_i(0) = 0$ and $D_i(1) = 1$).
- Always-takers: $D_i(1) = D_i(0) = 1$.
- Never-takers: $D_i(1) = D_i(0) = 0$.

Principal Strata

Following Angrist, Imbens, and Rubin (1996), we can define four subpopulations (for cases with a binary treatment and a binary instrument):

Definition

- Compliers: $D_i(1) > D_i(0)$ ($D_i(0) = 0$ and $D_i(1) = 1$).
- Always-takers: $D_i(1) = D_i(0) = 1$.
- Never-takers: $D_i(1) = D_i(0) = 0$.
- Defiers: $D_i(1) < D_i(0)$ ($D_i(0) = 1$ and $D_i(1) = 0$).

Principal Strata

Following Angrist, Imbens, and Rubin (1996), we can define four subpopulations (for cases with a binary treatment and a binary instrument):

Definition

- Compliers: $D_i(1) > D_i(0)$ ($D_i(0) = 0$ and $D_i(1) = 1$).
- Always-takers: $D_i(1) = D_i(0) = 1$.
- Never-takers: $D_i(1) = D_i(0) = 0$.
- Defiers: $D_i(1) < D_i(0)$ ($D_i(0) = 1$ and $D_i(1) = 0$).

Only one of the potential treatment indicators $(D_i(0), D_i(1))$ is observed, so in the general case we cannot identify exactly which group any particular individual belongs to (although we can rule some out).

Monotonicity means no defiers

Name	$D_i(1)$	$D_i(0)$
Always Takers	1	1
Never Takers	0	0
Compliers	1	0
Defiers	0	1

- We sometimes call assumption 4 no defiers because the monotonicity assumption rules out the existence of defiers.

Monotonicity means no defiers

Name	$D_i(1)$	$D_i(0)$
Always Takers	1	1
Never Takers	0	0
Compliers	1	0
Defiers	0	1

- We sometimes call assumption 4 no defiers because the monotonicity assumption rules out the existence of defiers.
- This means we can now sometimes identify the subgroup

Monotonicity means no defiers

Name	$D_i(1)$	$D_i(0)$
Always Takers	1	1
Never Takers	0	0
Compliers	1	0
Defiers	0	1

- We sometimes call assumption 4 no defiers because the monotonicity assumption rules out the existence of defiers.
- This means we can now sometimes identify the subgroup
- Anyone with $D_i = 1$ when $Z_i = 0$ must be an always-taker and anyone with $D_i = 0$ when $Z_i = 1$ must be a never-taker.

Local Average Treatment Effect (LATE)

- Under these four assumptions, we can use the Wald estimator to estimate the local average treatment effect (LATE) — sometimes called the complier average treatment effect.

Local Average Treatment Effect (LATE)

- Under these four assumptions, we can use the Wald estimator to estimate the local average treatment effect (LATE) — sometimes called the complier average treatment effect.
- This is the ATE among the compliers: those that take the treatment when encouraged to do so.

Local Average Treatment Effect (LATE)

- Under these four assumptions, we can use the Wald estimator to estimate the local average treatment effect (LATE) — sometimes called the complier average treatment effect.
- This is the ATE among the compliers: those that take the treatment when encouraged to do so.
- That is, the LATE theorem (proof in the appendix), states that:

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]$$

Local Average Treatment Effect (LATE)

- Under these four assumptions, we can use the Wald estimator to estimate the local average treatment effect (LATE) — sometimes called the complier average treatment effect.
- This is the ATE among the compliers: those that take the treatment when encouraged to do so.
- That is, the LATE theorem (proof in the appendix), states that:

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]$$

- This may seem mundane in that we have simply changed our assumptions and not our estimation, but this fact was a massive intellectual jump in our understanding of IV. Angrist, Imbens and Rubin (1996) is amazing, you should read it!

Who are the Compliers?

Study	Outcome	Treatment	Instrument
Angrist and Evans (1998)	Earnings	More than 2 Children	Multiple Second Birth (Twins)
Angrist and Evans (1998)	Earnings	More than 2 Children	First Two Children are Same Sex
Levitt (1997)	Crime Rates	Number of Policemen	Mayoral Elections
Angrist and Krueger (1991)	Earnings	Years of Schooling	Quarter of Birth
Angrist (1990)	Earnings	Veteran Status	Vietnam Draft Lottery
Miguel, Satyanath and Sergenti (2004)	Civil War Onset	GDP per capita	Lagged Rainfall
Acemoglu, Johnson and Robinson (2001)	Economic performance	Current Institutions	Settler Mortality in Colonial Times
Cleary and Barro (2006)	Religiosity	GDP per capita	Distance from Equator

Is the LATE useful?

- Once we allow for heterogeneous effects, all we can estimate with IV is the effect of treatment among compliers.

Is the LATE useful?

- Once we allow for heterogeneous effects, all we can estimate with IV is the effect of treatment among compliers.
- This is an unknown subset of the data.

Is the LATE useful?

- Once we allow for heterogeneous effects, all we can estimate with IV is the effect of treatment among compliers.
- This is an unknown subset of the data.
  - Treated units are a mix of always takers and compliers.
  - Control units are a mix of never takers and compliers.

Is the LATE useful?

- Once we allow for heterogeneous effects, all we can estimate with IV is the effect of treatment among compliers.
- This is an unknown subset of the data.
  - Treated units are a mix of always takers and compliers.
  - Control units are a mix of never takers and compliers.
- Without further assumptions, $\tau_{LATE} \neq \tau_{ATE}$.

Is the LATE useful?

- Once we allow for heterogeneous effects, all we can estimate with IV is the effect of treatment among compliers.
- This is an unknown subset of the data.
  - Treated units are a mix of always takers and compliers.
  - Control units are a mix of never takers and compliers.
- Without further assumptions, $\tau_{LATE} \neq \tau_{ATE}$.
- Complier group depends on the instrument $\rightsquigarrow$ different IVs will lead to different estimands!

Is the LATE useful?

- Once we allow for heterogeneous effects, all we can estimate with IV is the effect of treatment among compliers.
- This is an unknown subset of the data.
  - Treated units are a mix of always takers and compliers.
  - Control units are a mix of never takers and compliers.
- Without further assumptions, $\tau_{LATE} \neq \tau_{ATE}$.
- Complier group depends on the instrument $\rightsquigarrow$ different IVs will lead to different estimands!
- How much we care largely depends on our theory and what the instrument is.

Is the LATE useful?

- Once we allow for heterogeneous effects, all we can estimate with IV is the effect of treatment among compliers.
- This is an unknown subset of the data.
  - Treated units are a mix of always takers and compliers.
  - Control units are a mix of never takers and compliers.
- Without further assumptions, $\tau_{LATE} \neq \tau_{ATE}$.
- Complier group depends on the instrument $\rightsquigarrow$ different IVs will lead to different estimands!
- How much we care largely depends on our theory and what the instrument is.
- The traditional framework "cheats" by assuming that the effect is constant, so it is the same for compliers and non-compliers.

Randomized trials with one-sided noncompliance

- Will the LATE ever be equal to a usual causal quantity?

Randomized trials with one-sided noncompliance

- Will the LATE ever be equal to a usual causal quantity?
- When non-compliance is one-sided, then the LATE is equal to the ATT.

Randomized trials with one-sided noncompliance

- Will the LATE ever be equal to a usual causal quantity?
- When non-compliance is one-sided, then the LATE is equal to the ATT.
- Think of a randomized experiment:
  - Randomized treatment assignment = instrument (Z_i)
  - Non-randomized actual treatment taken = treatment (D_i)

Randomized trials with one-sided noncompliance

- Will the LATE ever be equal to a usual causal quantity?
- When non-compliance is one-sided, then the LATE is equal to the ATT.
- Think of a randomized experiment:
  - Randomized treatment assignment = instrument (Z_i)
  - Non-randomized actual treatment taken = treatment (D_i)
- One-sided noncompliance: only those assigned to treatment (control) can actually take the treatment (control). Or

$$D_i(0) = 0 \forall i \quad \rightsquigarrow \quad \Pr[D_i = 1 | Z_i = 0] = 0$$

Randomized trials with one-sided noncompliance

- Will the LATE ever be equal to a usual causal quantity?
- When non-compliance is one-sided, then the LATE is equal to the ATT.
- Think of a randomized experiment:
  - Randomized treatment assignment = instrument (Z_i)
  - Non-randomized actual treatment taken = treatment (D_i)
- One-sided noncompliance: only those assigned to treatment (control) can actually take the treatment (control). Or

$$D_i(0) = 0 \forall i \quad \rightsquigarrow \quad \Pr[D_i = 1 | Z_i = 0] = 0$$

- Maybe this is because only those treated actually get pills or only they are invited to the job training location.

Randomized trials with one-sided noncompliance

- Will the LATE ever be equal to a usual causal quantity?
- When non-compliance is one-sided, then the LATE is equal to the ATT.
- Think of a randomized experiment:
  - Randomized treatment assignment = instrument (Z_i)
  - Non-randomized actual treatment taken = treatment (D_i)
- One-sided noncompliance: only those assigned to treatment (control) can actually take the treatment (control). Or

$$D_i(0) = 0 \forall i \quad \rightsquigarrow \quad \Pr[D_i = 1 | Z_i = 0] = 0$$

- Maybe this is because only those treated actually get pills or only they are invited to the job training location.
- Note: this can be very difficult to do practically in many settings.

Benefits of one-sided noncompliance

One-sided noncompliance $\rightsquigarrow$ no "always-takers" and since there are no defiers,

- Treated units must be compliers.
- ATT is the same as the LATE.

Benefits of one-sided noncompliance

One-sided noncompliance $\rightsquigarrow$ no "always-takers" and since there are no defiers,

- Treated units must be compliers.
- ATT is the same as the LATE.

Proof.

$$
\begin{aligned}
E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =& \mathbb{E}[Y_i(0) + (Y_i(1) - Y_i(0))D_i|Z_i = 1] - \mathbb{E}[Y_i(0)|Z_i = 0] \\
& \text{(exclusion restriction + one-sided noncompliance)} \\
=& \mathbb{E}[Y_i(0)|Z_i = 1] + E[(Y_i(1) - Y_i(0))D_i|Z_i = 1] - \mathbb{E}[Y_i(0)|Z_i = 0] \\
=& \mathbb{E}[Y_i(0)] + \mathbb{E}[(Y_i(1) - Y_i(0))D_i|Z_i = 1] - \mathbb{E}[Y_i(0)] \\
& \text{(randomization)} \\
=& \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1, Z_i = 1] \Pr[D_i = 1|Z_i = 1] \\
& \text{(law of iterated expectations + binary treatment)} \\
=& \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1] \Pr[D_i = 1|Z_i = 1] \\
& \text{(one-sided noncompliance)}
\end{aligned}
$$

Noting that $\Pr[D_i = 1|Z_i = 0] = 0$, then the Wald estimator is just the ATT: $\frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{\Pr[D_i=1|Z_i=1]} = E[Y_i(1) - Y_i(0)|D_i = 1]$ Thus, under the additional assumption of one-sided compliance, we can estimate the ATT using the usual IV approach □

Example: The Vietnam Draft Lottery (Angrist (1990))

Example: The Vietnam Draft Lottery (Angrist (1990))

- Effect of military service on civilian earnings

Example: The Vietnam Draft Lottery (Angrist (1990))

- Effect of military service on civilian earnings
- Simple comparison between Vietnam veterans and non-veterans are likely to be a biased measure

Example: The Vietnam Draft Lottery (Angrist (1990))

- Effect of military service on civilian earnings
- Simple comparison between Vietnam veterans and non-veterans are likely to be a biased measure
- Angrist (1990) used draft-eligibility, determined by the Vietnam era draft lottery, as an instrument for military service in Vietnam

Example: The Vietnam Draft Lottery (Angrist (1990))

- Effect of military service on civilian earnings
- Simple comparison between Vietnam veterans and non-veterans are likely to be a biased measure
- Angrist (1990) used draft-eligibility, determined by the Vietnam era draft lottery, as an instrument for military service in Vietnam
- Draft eligibility is random and affected the probability of enrollment

Example: The Vietnam Draft Lottery (Angrist (1990))

- Effect of military service on civilian earnings
- Simple comparison between Vietnam veterans and non-veterans are likely to be a biased measure
- Angrist (1990) used draft-eligibility, determined by the Vietnam era draft lottery, as an instrument for military service in Vietnam
- Draft eligibility is random and affected the probability of enrollment
- Estimate suggest a 15% negative effect of veteran status on earnings in the period 1981-1984 for white veterans born in 1950-51; although the estimators are quite imprecise

Example: The Vietnam Draft Lottery (Angrist (1990))

- Effect of military service on civilian earnings
- Simple comparison between Vietnam veterans and non-veterans are likely to be a biased measure
- Angrist (1990) used draft-eligibility, determined by the Vietnam era draft lottery, as an instrument for military service in Vietnam
- Draft eligibility is random and affected the probability of enrollment
- Estimate suggest a 15% negative effect of veteran status on earnings in the period 1981-1984 for white veterans born in 1950-51; although the estimators are quite imprecise
- This is only identified for compliers (i.e. those who if draft eligible would serve but otherwise would not)

Wald Estimates for Vietnam Draft Lottery (Angrist (1990))

		Draft-Eligibility Effects in Current $				
		FICA Earnings	Adjusted FICA Earnings	Total W-2 Earnings	$\hat{p}^e - \hat{p}^n$	Service Effect in 1978 $
Cohort	Year	(1)	(2)	(3)	(4)	(5)
1950	1981	−435.8	−487.8	−589.6	0.159	−2,195.8
		(210.5)	(237.6)	(299.4)	(0.040)	(1,069.5)
	1982	−320.2	−396.1	−305.5		−1,678.3
		(235.8)	(281.7)	(345.4)		(1,193.6)
	1983	−349.5	−450.1	−512.9		−1,795.6
		(261.6)	(302.0)	(441.2)		(1,204.8)
	1984	−484.3	−638.7	−1,143.3		−2,517.7
		(286.8)	(336.5)	(492.2)		(1,326.5)
1951	1981	−358.3	−428.7	−71.6	0.136	−2,261.3
		(203.6)	(224.5)	(423.4)	(0.043)	(1,184.2)
	1982	−117.3	−278.5	−72.7		−1,386.6
		(229.1)	(264.1)	(372.1)		(1,312.1)
	1983	−314.0	−452.2	−896.5		-2,181.8
		(253.2)	(289.2)	(426.3)		(1,395.3)
	1984	−398.4	−573.3	−809.1		−2,647.9
		(279.2)	(331.1)	(380.9)		(1,529.2)
1952	1981	−342.8	−392.6	−440.5	0.105	−2,502.3
		(206.8)	(228.6)	(265.0)	(0.050)	(1,556.7)
	1982	−235.1	−255.2	−514.7		−1,626.5
		(232.3)	(264.5)	(296.5)		(1,685.8)
	1983	−437.7	−500.0	−915.7		−3,103.5
		(257.5)	(294.7)	(395.2)		(1,829.2)
	1984	−436.0	−560.0	−767.2		−3,323.8
		(281.9)	(330.1)	(376.0)		(1,959.3)

Estimating the Size of the Complier Group

Estimating the Size of the Complier Group

- Since we never observe both potential treatment assignments for the same unit, we cannot identify individual units as compliers

Estimating the Size of the Complier Group

- Since we never observe both potential treatment assignments for the same unit, we cannot identify individual units as compliers
- However, we can easily identify the proportion of compliers in the population using the first stage effect:

Estimating the Size of the Complier Group

- Since we never observe both potential treatment assignments for the same unit, we cannot identify individual units as compliers
- However, we can easily identify the proportion of compliers in the population using the first stage effect:

$$
\begin{aligned}
P(D_i(1) > D_i(0)) &= E[D_i(1) - D_i(0)] = E[D_i(1)] - E[D_i(0)] \\
&= E[D_i|Z_i = 1] - E[D_i|Z_i = 0]
\end{aligned}
$$

Estimating the Size of the Complier Group

- Since we never observe both potential treatment assignments for the same unit, we cannot identify individual units as compliers
- However, we can easily identify the proportion of compliers in the population using the first stage effect:

$$\begin{aligned} P(D_i(1) > D_i(0)) &= E[D_i(1) - D_i(0)] = E[D_i(1)] - E[D_i(0)] \\ &= E[D_i|Z_i = 1] - E[D_i|Z_i = 0] \end{aligned}$$

- Using a similar logic we can identify the proportion of compliers among the treated or controls only.

Estimating the Size of the Complier Group

- Since we never observe both potential treatment assignments for the same unit, we cannot identify individual units as compliers
- However, we can easily identify the proportion of compliers in the population using the first stage effect:

$$\begin{aligned} P(D_i(1) > D_i(0)) &= E[D_i(1) - D_i(0)] = E[D_i(1)] - E[D_i(0)] \\ &= E[D_i|Z_i = 1] - E[D_i|Z_i = 0] \end{aligned}$$

- Using a similar logic we can identify the proportion of compliers among the treated or controls only. For example:

$$P(D_i(1) > D_i(0)|D_i = 1) = \frac{P(Z_i = 1)(E[D_i|Z_i = 1] - E[D_i|Z_i = 0])}{P(D_i = 1)}$$

Estimating the Size of the Complier Group

- Since we never observe both potential treatment assignments for the same unit, we cannot identify individual units as compliers
- However, we can easily identify the proportion of compliers in the population using the first stage effect:

$$\begin{aligned} P(D_i(1) > D_i(0)) &= E[D_i(1) - D_i(0)] = E[D_i(1)] - E[D_i(0)] \\ &= E[D_i|Z_i = 1] - E[D_i|Z_i = 0] \end{aligned}$$

- Using a similar logic we can identify the proportion of compliers among the treated or controls only. For example:

$$P(D_i(1) > D_i(0)|D_i = 1) = \frac{P(Z_i = 1)(E[D_i|Z_i = 1] - E[D_i|Z_i = 0])}{P(D_i = 1)}$$

- Note: this estimate is pinned down entirely by the assumptions of monotonicity and exogeneity

Estimating the Size of the Complier Group

- Since we never observe both potential treatment assignments for the same unit, we cannot identify individual units as compliers
- However, we can easily identify the proportion of compliers in the population using the first stage effect:

$$\begin{aligned} P(D_i(1) > D_i(0)) &= E[D_i(1) - D_i(0)] = E[D_i(1)] - E[D_i(0)] \\ &= E[D_i|Z_i = 1] - E[D_i|Z_i = 0] \end{aligned}$$

- Using a similar logic we can identify the proportion of compliers among the treated or controls only. For example:

$$P(D_i(1) > D_i(0)|D_i = 1) = \frac{P(Z_i = 1)(E[D_i|Z_i = 1] - E[D_i|Z_i = 0])}{P(D_i = 1)}$$

- Note: this estimate is pinned down entirely by the assumptions of monotonicity and exogeneity
- Abadie (2003) shows how to use covariate information to calculate other characteristics of the complier group (kappa weighting)

Size of Complier Group

TABLE 4.4.2
Probabilities of compliance in instrumental variables studies

Source (1)	Endogenous Variable (D) (2)	Instrument (Z) (3)	Sample (4)	$P[D=1]$ (5)	First Stage, $P[D_1 > D_0]$ (6)	$P[Z=1]$ (7)	Compliance Probabilities $P[D_1 > D_0 \mid D=1]$ (8)	$P[D_1 > D_0 \mid D=0]$ (9)
Angrist (1990)	Veteran status	Draft eligibility	White men born in 1950	.267	.159	.534	.318	.101
			Non-white men born in 1950	.163	.060	.534	.197	.033
Angrist and Evans (1998)	More than two children	Twins at second birth	Married women aged 21–35 with two or more children in 1980	.381	.603	.008	.013	.966
		First two children are same sex		.381	.060	.506	.080	.048
Angrist and Krueger (1991)	High school graduate	Third- or fourth-quarter birth	Men born between 1930 and 1939	.770	.016	.509	.011	.034
Acemoglu and Angrist (2000)	High school graduate	State requires 11 or more years of school attendance	White men aged 40–49	.617	.037	.300	.018	.068

Notes: The table computes the absolute and relative size of the complier population for a number of instrumental variables. The first stage, reported in column 6, gives the absolute size of the complier group. Columns 8 and 9 show the size of the complier population relative to the treated and untreated populations.

Do we care about the complier effect?

IV estimates the effect for compliers. How do we feel about that?

Do we care about the complier effect?

IV estimates the effect for compliers. How do we feel about that?

- Pros
  - it is identified (i.e. it is what we can get)

Do we care about the complier effect?

IV estimates the effect for compliers. How do we feel about that?

- Pros
  - it is identified (i.e. it is what we can get)
  - if the instrument is the policy lever, these are the people we can help

Do we care about the complier effect?

IV estimates the effect for compliers. How do we feel about that?

- Pros
  - it is identified (i.e. it is what we can get)
  - if the instrument is the policy lever, these are the people we can help
  - we can compute proportion of compliers (and under strong assumptions) characterize the compliers in terms of observed variables

Do we care about the complier effect?

IV estimates the effect for compliers. How do we feel about that?

- Pros
  - it is identified (i.e. it is what we can get)
  - if the instrument is the policy lever, these are the people we can help
  - we can compute proportion of compliers (and under strong assumptions) characterize the compliers in terms of observed variables
- Cons
  - we don't actually know who the compliers are and every instrument yields a different group

Do we care about the complier effect?

IV estimates the effect for compliers. How do we feel about that?

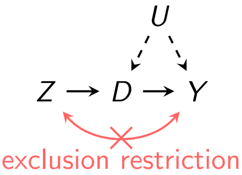
- Pros
  - it is identified (i.e. it is what we can get)
  - if the instrument is the policy lever, these are the people we can help
  - we can compute proportion of compliers (and under strong assumptions) characterize the compliers in terms of observed variables
- Cons
  - we don't actually know who the compliers are and every instrument yields a different group
  - monotonicity is a strong unit-level assumption
    (i.e. it is unlikely to hold when decision to treat is the result of multiple criteria that includes risks and benefits, see Hernán and Robins 2018, pg 63)

Do we care about the complier effect?

IV estimates the effect for compliers. How do we feel about that?

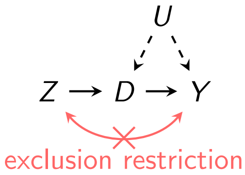
- Pros
  - it is identified (i.e. it is what we can get)
  - if the instrument is the policy lever, these are the people we can help
  - we can compute proportion of compliers (and under strong assumptions) characterize the compliers in terms of observed variables
- Cons
  - we don't actually know who the compliers are and every instrument yields a different group
  - monotonicity is a strong unit-level assumption
    (i.e. it is unlikely to hold when decision to treat is the result of multiple criteria that includes risks and benefits, see Hernán and Robins 2018, pg 63)
  - 'relatively minor violations of conditions [Assumptions 1-4] for IV estimation may result in large biases of unpredictable or counter-intuitive direction'
    (Hernán and Robins 2018)

Falsification tests



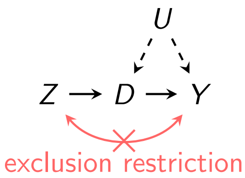
- The exclusion restriction cannot be tested directly, but it can be falsified.

Falsification tests



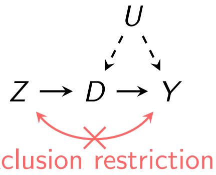
- The exclusion restriction cannot be tested directly, but it can be falsified.
- Falsification test Test the reduced form effect of Z_i on Y_i in situations where it is impossible or extremely unlikely that Z_i could affect D_i.

Falsification tests



- The exclusion restriction cannot be tested directly, but it can be falsified.
- Falsification test Test the reduced form effect of Z_i on Y_i in situations where it is impossible or extremely unlikely that Z_i could affect D_i.
- Because Z_i can't affect D_i, then the exclusion restriction implies that this falsification test should have 0 effect.

Falsification tests



- The exclusion restriction cannot be tested directly, but it can be falsified.
- Falsification test Test the reduced form effect of Z_i on Y_i in situations where it is impossible or extremely unlikely that Z_i could affect D_i.
- Because Z_i can't affect D_i, then the exclusion restriction implies that this falsification test should have 0 effect.
- Nunn & Wantchekon (2011): use distance to coast as an instrument for Africans, use distance to the coast in an Asian sample as falsification test.

Nunn & Wantchekon falsification test

TABLE 7—REDUCED FORM RELATIONSHIP BETWEEN THE DISTANCE FROM THE COAST AND TRUST WITHIN AFRICA AND ASIA

	Trust of local government council			
	Afrobarometer sample		Asiabarometer sample	
	(1)	(2)	(3)	(4)
Distance from the coast	0.00039***	0.00031***	−0.00001	0.00001
	(0.00009)	(0.00008)	(0.00010)	(0.00009)
Country fixed effects	Yes	Yes	Yes	Yes
Individual controls	No	Yes	No	Yes
Number of observations	19,913	19,913	5,409	5,409
Number of clusters	185	185	62	62
R^2	0.16	0.18	0.19	0.22

Notes: The table reports OLS estimates. The unit of observation is an individual. The dependent variable in the Asiabarometer sample is the respondent's answer to the question: "How much do you trust your local government?" The categories for the answers are the same in the Asiabarometer as in the Afrobarometer. Standard errors are clustered at the ethnicity level in the Afrobarometer regressions and at the location (city) level in the Asiabarometer and the WVS samples. The individual controls are for age, age squared, a gender indicator, education fixed effects, and religion fixed effects.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Other Extensions to IV

- Multiple instruments

Other Extensions to IV

- Multiple instruments
- Covariates and conditional ignorability (see Glynn and Rueda 2018 on post-instrument bias though)

Other Extensions to IV

- Multiple instruments
- Covariates and conditional ignorability (see Glynn and Rueda 2018 on post-instrument bias though)
- Overidentification tests (in constant effects)

Other Extensions to IV

- Multiple instruments
- Covariates and conditional ignorability (see Glynn and Rueda 2018 on post-instrument bias though)
- Overidentification tests (in constant effects)
- Compliance modeling for weak instruments

Other Extensions to IV

- Multiple instruments
- Covariates and conditional ignorability (see Glynn and Rueda 2018 on post-instrument bias though)
- Overidentification tests (in constant effects)
- Compliance modeling for weak instruments
- Conditional effects and causal interaction models

Classical Vs. Modern Instrumental Variables

Classical Vs. Modern Instrumental Variables

- We dropped the constant effects assumption (assumption 4), which is usually unrealistic.

Classical Vs. Modern Instrumental Variables

- We dropped the constant effects assumption (assumption 4), which is usually unrealistic.
- We added a weaker monotonicity assumption.

Classical Vs. Modern Instrumental Variables

- We dropped the constant effects assumption (assumption 4), which is usually unrealistic.
- We added a weaker monotonicity assumption.
- We defined a set of subpopulations: compilers, always-takers, never-takers, defiers

Classical Vs. Modern Instrumental Variables

- We dropped the constant effects assumption (assumption 4), which is usually unrealistic.
- We added a weaker monotonicity assumption.
- We defined a set of subpopulations: compilers, always-takers, never-takers, defiers
- We clarify that the effects are identified only for a particular subpopulation — the "complier" subpopulation.

Classical Vs. Modern Instrumental Variables

- We dropped the constant effects assumption (assumption 4), which is usually unrealistic.
- We added a weaker monotonicity assumption.
- We defined a set of subpopulations: compilers, always-takers, never-takers, defiers
- We clarify that the effects are identified only for a particular subpopulation — the "complier" subpopulation.
  (if constant effects happen to hold, effects for compliers are by definition same as for entire population.)

Wait, What Were The Assumptions?

Wait, What Were The Assumptions?

- Assumptions given have typically been sufficient but not necessary.

Wait, What Were The Assumptions?

- Assumptions given have typically been sufficient but not necessary.
- Easy to state assumptions for linear structural equation models (where we can use covariance of error and variable), harder in general.

Wait, What Were The Assumptions?

- Assumptions given have typically been sufficient but not necessary.
- Easy to state assumptions for linear structural equation models (where we can use covariance of error and variable), harder in general.
- Technically can use the following in place of assumptions 1-3: (see technical point 16.1 of Hernán and Robins)

Wait, What Were The Assumptions?

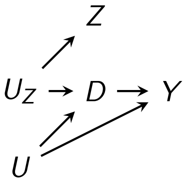
- Assumptions given have typically been sufficient but not necessary.
- Easy to state assumptions for linear structural equation models (where we can use covariance of error and variable), harder in general.
- Technically can use the following in place of assumptions 1-3: (see technical point 16.1 of Hernán and Robins)
  1) Exogeneity: $Y_i(d,z) \perp\!\!\!\perp Z_i$ for all d, z.
  2) Exclusion: $Y_i(d,z) = Y_i(d,z') = Y_i(d)$ for all z, z', d and i
  3) Relevance: $Z \not\perp\!\!\!\perp D$
  - all possibly conditional on X

Wait, What Were The Assumptions?

- Assumptions given have typically been sufficient but not necessary.
- Easy to state assumptions for linear structural equation models (where we can use covariance of error and variable), harder in general.
- Technically can use the following in place of assumptions 1-3: (see technical point 16.1 of Hernán and Robins)
  1) Exogeneity: $Y_i(d,z) \perp\!\!\!\perp Z_i$ for all d, z.
  2) Exclusion: $Y_i(d,z) = Y_i(d,z') = Y_i(d)$ for all z, z', d and i
  3) Relevance: $Z \not\perp\!\!\!\perp D$
     - all possibly conditional on X
- This allows some graphs that don't meet our original conditions, but satisfy new assumptions given particular edge configurations.

Wait, What Were The Assumptions?

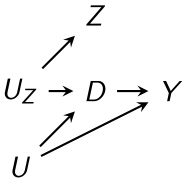
- Assumptions given have typically been sufficient but not necessary.
- Easy to state assumptions for linear structural equation models (where we can use covariance of error and variable), harder in general.
- Technically can use the following in place of assumptions 1-3: (see technical point 16.1 of Hernán and Robins)
  1) Exogeneity: $Y_i(d, z) \perp\!\!\!\perp Z_i$ for all d, z.
  2) Exclusion: $Y_i(d, z) = Y_i(d, z') = Y_i(d)$ for all z, z', d and i
  3) Relevance: $Z \not\!\perp\!\!\!\perp D$
     - all possibly conditional on X
- This allows some graphs that don't meet our original conditions, but satisfy new assumptions given particular edge configurations.



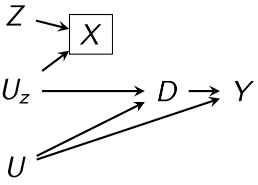
(proxy instrument)

Wait, What Were The Assumptions?

- Assumptions given have typically been sufficient but not necessary.
- Easy to state assumptions for linear structural equation models (where we can use covariance of error and variable), harder in general.
- Technically can use the following in place of assumptions 1-3: (see technical point 16.1 of Hernán and Robins)
  1) Exogeneity: $Y_i(d,z) \perp\!\!\!\perp Z_i$ for all d, z.
  2) Exclusion: $Y_i(d,z) = Y_i(d,z') = Y_i(d)$ for all z, z', d and i
  3) Relevance: $Z \not\perp\!\!\!\perp D$
     - all possibly conditional on X
- This allows some graphs that don't meet our original conditions, but satisfy new assumptions given particular edge configurations.



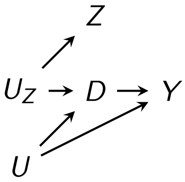
(proxy instrument)



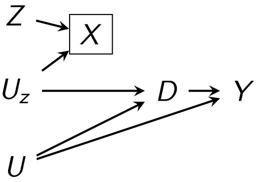
(instrument via collider)

Wait, What Were The Assumptions?

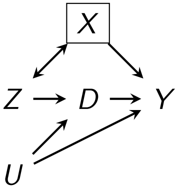
- Assumptions given have typically been sufficient but not necessary.
- Easy to state assumptions for linear structural equation models (where we can use covariance of error and variable), harder in general.
- Technically can use the following in place of assumptions 1-3: (see technical point 16.1 of Hernán and Robins)
  1) Exogeneity: $Y_i(d,z) \perp\!\!\!\perp Z_i$ for all d, z.
  2) Exclusion: $Y_i(d,z) = Y_i(d,z') = Y_i(d)$ for all z, z', d and i
  3) Relevance: $Z \not\perp\!\!\!\perp D$
     - all possibly conditional on X
- This allows some graphs that don't meet our original conditions, but satisfy new assumptions given particular edge configurations.



(proxy instrument)



(instrument via collider)



(exclusion via control)

Concluding Thoughts on Instrumental Variables

Concluding Thoughts on Instrumental Variables

- Strong assumptions but powerful results

Concluding Thoughts on Instrumental Variables

- Strong assumptions but powerful results
- Enormous care is required in the interpretation

Concluding Thoughts on Instrumental Variables

- Strong assumptions but powerful results
- Enormous care is required in the interpretation
- Questions to Always Ask

Concluding Thoughts on Instrumental Variables

- Strong assumptions but powerful results
- Enormous care is required in the interpretation
- Questions to Always Ask
    1. is the instrument weak?
       (does Z predict D)

Concluding Thoughts on Instrumental Variables

- Strong assumptions but powerful results
- Enormous care is required in the interpretation
- Questions to Always Ask
  1. is the instrument weak?
     (does Z predict D)
  2. is the instrument exogenous?
     (was Z randomly assigned?)

Concluding Thoughts on Instrumental Variables

- Strong assumptions but powerful results
- Enormous care is required in the interpretation
- Questions to Always Ask
  1. is the instrument weak?
     (does Z predict D)
  2. is the instrument exogenous?
     (was Z randomly assigned?)
  3. does the exclusion restriction hold?
     (is there a path from Z to Y not through D)

Concluding Thoughts on Instrumental Variables

- Strong assumptions but powerful results
- Enormous care is required in the interpretation
- Questions to Always Ask
  1. is the instrument weak?
     (does Z predict D)
  2. is the instrument exogenous?
     (was Z randomly assigned?)
  3. does the exclusion restriction hold?
     (is there a path from Z to Y not through D)
  4. do we believe monotonicity?
     (are there units where the instrument discourages treatment)

Concluding Thoughts on Instrumental Variables

- Strong assumptions but powerful results
- Enormous care is required in the interpretation
- Questions to Always Ask
  1. is the instrument weak?
     (does Z predict D)
  2. is the instrument exogenous?
     (was Z randomly assigned?)
  3. does the exclusion restriction hold?
     (is there a path from Z to Y not through D)
  4. do we believe monotonicity?
     (are there units where the instrument discourages treatment)
  5. do the assumptions identify an effect for the subpopulation of interest?
     (is the instrument such that we care about compliers?)

Concluding Thoughts on Instrumental Variables

- Strong assumptions but powerful results
- Enormous care is required in the interpretation
- Questions to Always Ask
  1. is the instrument weak?
     (does Z predict D)
  2. is the instrument exogenous?
     (was Z randomly assigned?)
  3. does the exclusion restriction hold?
     (is there a path from Z to Y not through D)
  4. do we believe monotonicity?
     (are there units where the instrument discourages treatment)
  5. do the assumptions identify an effect for the subpopulation of interest?
     (is the instrument such that we care about compliers?)
- Be sure to evaluate all conditions and remember randomization of Z does not guarantee the exclusion restriction.

Regression Discontinuity

- A different strategy where the core intuition is that identification comes in a discontinuity in treatment assignment

Regression Discontinuity

- A different strategy where the core intuition is that identification comes in a discontinuity in treatment assignment
- A widely applicable strategy in rule-based systems or allocations of limited resources (e.g. administrative programs, elections, admission systems)

Regression Discontinuity

- A different strategy where the core intuition is that identification comes in a discontinuity in treatment assignment
- A widely applicable strategy in rule-based systems or allocations of limited resources (e.g. administrative programs, elections, admission systems)
- It is a fairly old idea, generally credited to education research by Thistlethwaite and Campbell 1960 but with a dynamic and interesting recent history (Hahn et al 2001 and Lee 2008 were big jumps forward).

Regression Discontinuity

- A different strategy where the core intuition is that identification comes in a discontinuity in treatment assignment
- A widely applicable strategy in rule-based systems or allocations of limited resources (e.g. administrative programs, elections, admission systems)
- It is a fairly old idea, generally credited to education research by Thistlethwaite and Campbell 1960 but with a dynamic and interesting recent history (Hahn et al 2001 and Lee 2008 were big jumps forward).
- The goal here is to get you up to speed with the core idea: if you want to know how to do this in practice read *A Practical Introduction to Regression Discontinuity Designs* Volumes I and II by Matias Cattaneo, Nicolás Idrodo and Rocío Titiunik

Setup

- The basic idea behind RDDs:

Setup

- The basic idea behind RDDs:
  - X_i is a forcing variable.

Setup

- The basic idea behind RDDs:
  - X_i is a forcing variable.
  - Treatment assignment is determined by a cutoff in X_i.

$$D_i = 1\{X_i > c\} \text{ so } D_i = \begin{cases} D_i = 1 & \text{if } X_i > c \\ D_i = 0 & \text{if } X_i < c \end{cases}$$

Setup

- The basic idea behind RDDs:
  - X_i is a forcing variable.
  - Treatment assignment is determined by a cutoff in X_i.

$$D_i = 1\{X_i > c\} \text{ so } D_i = \begin{cases} D_i = 1 & \text{if } X_i > c \\ D_i = 0 & \text{if } X_i < c \end{cases}$$

- X_i can be related to the potential outcomes and so comparing treated and untreated units does not provide causal estimates

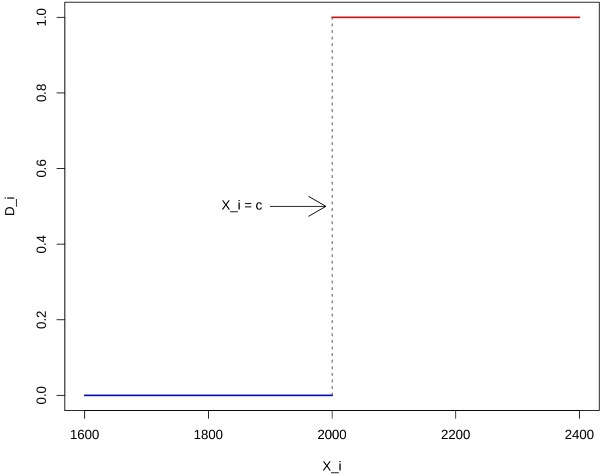
Setup

- The basic idea behind RDDs:
  - X_i is a forcing variable.
  - Treatment assignment is determined by a cutoff in X_i.

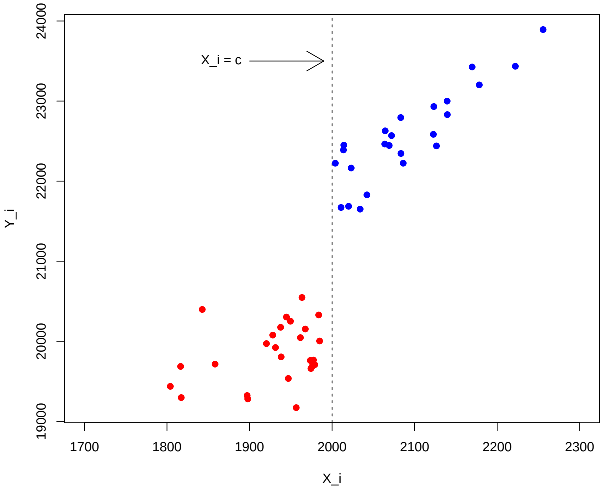
$$D_i = 1\{X_i > c\} \text{ so } D_i = \begin{cases} D_i = 1 & \text{if } X_i > c \\ D_i = 0 & \text{if } X_i < c \end{cases}$$

- X_i can be related to the potential outcomes and so comparing treated and untreated units does not provide causal estimates
- assume relationship between X and the potential outcomes Y_1 and Y_0 is smooth around the threshold $\rightsquigarrow$ discontinuity created by the treatment to estimate the effect of D on Y at the threshold

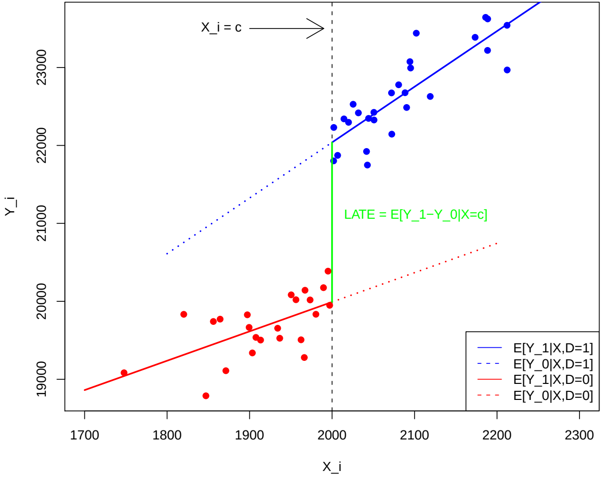
Graphical Illustration



Graphical Illustration



Graphical Illustration



Design

- Sharp RD: treatment assignment is a deterministic function of the forcing variable and the threshold.

Design

- Sharp RD: treatment assignment is a deterministic function of the forcing variable and the threshold.
- Key assumption: no compliance problems (deterministic)

Design

- Sharp RD: treatment assignment is a deterministic function of the forcing variable and the threshold.
- Key assumption: no compliance problems (deterministic)
- At the threshold, c, we only see treated units and below the threshold $c - \varepsilon$, we only see control values:

$$\mathbb{P}(D_i = 1 | X_i = c) = 1$$
$$\mathbb{P}(D_i = 1 | X_i = c - \varepsilon) = 0$$

Design

- Sharp RD: treatment assignment is a deterministic function of the forcing variable and the threshold.
- Key assumption: no compliance problems (deterministic)
- At the threshold, c, we only see treated units and below the threshold $c - \varepsilon$, we only see control values:

$$\mathbb{P}(D_i = 1 | X_i = c) = 1$$
$$\mathbb{P}(D_i = 1 | X_i = c - \varepsilon) = 0$$

- Intuitively, we are interested in the discontinuity in the outcome at the discontinuity in the treatment assignment.

Design

- Sharp RD: treatment assignment is a deterministic function of the forcing variable and the threshold.
- Key assumption: no compliance problems (deterministic)
- At the threshold, c, we only see treated units and below the threshold $c - \varepsilon$, we only see control values:

$$\mathbb{P}(D_i = 1 | X_i = c) = 1$$
$$\mathbb{P}(D_i = 1 | X_i = c - \varepsilon) = 0$$

- Intuitively, we are interested in the discontinuity in the outcome at the discontinuity in the treatment assignment.
- We want to investigate the behavior of the outcome around the threshold: $\lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x]$

Design

- Sharp RD: treatment assignment is a deterministic function of the forcing variable and the threshold.
- Key assumption: no compliance problems (deterministic)
- At the threshold, c, we only see treated units and below the threshold $c - \varepsilon$, we only see control values:

$$\mathbb{P}(D_i = 1 | X_i = c) = 1$$
$$\mathbb{P}(D_i = 1 | X_i = c - \varepsilon) = 0$$

- Intuitively, we are interested in the discontinuity in the outcome at the discontinuity in the treatment assignment.
- We want to investigate the behavior of the outcome around the threshold: $\lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x]$
- Under certain assumptions, this quantity identifies the ATE at the threshold: $\tau_{SRD} = E[Y_i(1) - Y_i(0) | X_i = c]$

Identification

Identification Assumption

1. $Y_1, Y_0 \perp\!\!\!\perp D | X$ *(trivially met by construction)*
2. $0 < P(D = 1 | X = x) < 1$ *(always violated in Sharp RDD)*
3. $E[Y_1|X, D]$ *and* $E[Y_0|X, D]$ *are continuous in* X *around the threshold* $X = c$ *(individuals have imprecise control over* X *around the threshold)*

Identification Result

The treatment effect is identified at the threshold as:

$$\begin{aligned} \alpha_{SRDD} &= E[Y_1 - Y_0 | X = c] \\ &= E[Y_1|X = c] - E[Y_0|X = c] \\ &= \lim_{x \downarrow c} E[Y_1|X = x] - \lim_{x \uparrow c} E[Y_0|X = x] \end{aligned}$$

Without further assumptions α_{SRDD} *is only identified at the threshold.*

Extrapolation and smoothness

- Remember the quantity of interest here is the effect at the threshold:

$$\begin{aligned}\tau_{SRD} &= E[Y_i(1) - Y_i(0)|X_i = c] \\ &= E[Y_i(1)|X_i = c] - E[Y_i(0)|X_i = c]\end{aligned}$$

Extrapolation and smoothness

- Remember the quantity of interest here is the effect at the threshold:

$$\begin{aligned}\tau_{SRD} &= E[Y_i(1) - Y_i(0)|X_i = c] \\ &= E[Y_i(1)|X_i = c] - E[Y_i(0)|X_i = c]\end{aligned}$$

- But we don't observe $E[Y_i(0)|X_i = c]$ ever due to the design, so we're going to extrapolate from $E[Y_i(0)|X_i = c - \varepsilon]$.

Extrapolation and smoothness

- Remember the quantity of interest here is the effect at the threshold:

$$\begin{aligned}\tau_{SRD} &= E[Y_i(1) - Y_i(0)|X_i = c] \\ &= E[Y_i(1)|X_i = c] - E[Y_i(0)|X_i = c]\end{aligned}$$

- But we don't observe $E[Y_i(0)|X_i = c]$ ever due to the design, so we're going to extrapolate from $E[Y_i(0)|X_i = c - \varepsilon]$.
- Extrapolation, even at short distances, requires smoothness in the functions we are extrapolating.

What can go wrong?

What can go wrong?

- If the potential outcomes change at the discontinuity for reasons other than the treatment, then smoothness will be violated.

What can go wrong?

- If the potential outcomes change at the discontinuity for reasons other than the treatment, then smoothness will be violated.
- For instance, if people sort around threshold, then you might get jumps other than the one you care about.

What can go wrong?

- If the potential outcomes change at the discontinuity for reasons other than the treatment, then smoothness will be violated.
- For instance, if people sort around threshold, then you might get jumps other than the one you care about.
- If things other than the treatment change at the threshold, then that might cause discontinuities in the potential outcomes.

Example: Electronic Voting (Hidalgo 2012)

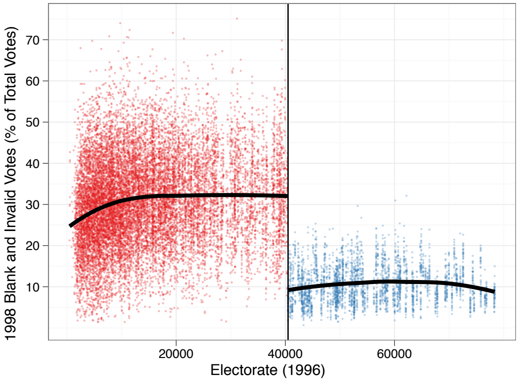


Figure 6: The effect of electronic voting on the percent of null and blank votes. Each dot is a polling station. Polling stations to the left of the vertical black line used paper ballots and polling stations to the right used electronic voting. The black horizontal line is the conditional mean of the outcome estimated with a loess regression.

Other Recent RDD Examples

- class size on student achievement
  - Angrist and Lavy 1999
- wage increase on performance of mayors
  Ferraz and Finan 2011; Gagliarducci and Nannicini 2013
- colonial institutions on development outcomes
  Dell 2009
- length of postpartum hospital stays on mother and infant mortality
  Almond and Doyle 2009
- naturalization on political integration of immigrants
  Hainmueller and Hangartner 2015
- financial aid offers on college enrollment
  Van der Klaauw 2002
- access to Angel funding on growth of start-ups
  Kerr, Lerner and Schoar 2010
- RDD that exploits "close" elections is workhorse model for electoral research:
  Lee, Moretti and Butler 2004, DiNardo and Lee 2004, Hainmueller and Kern 2008, Leigh 2008, Pettersson-Lidbom 2008, Broockman 2009, Butler 2009, Dal Bó, Dal Bó and Snyder 2009, Eggers and Hainmueller 2009, Ferreira and Gyourko 2009, Uppal 2009, 2010, Cellini, Ferreira and Rothstein 2010, Gerber and Hopkins 2011, Trounstine 2011, Boas and Hidalgo 2011, Folke and Snyder Jr. 2012, and Gagliarducci and Paserman 2012

General estimation strategy

- The main goal in RD is to estimate the limits of various CEFs such as:

$$\lim_{x \uparrow c} E[Y_i | X_i = x]$$

General estimation strategy

- The main goal in RD is to estimate the limits of various CEFs such as:

$$\lim_{x \uparrow c} E[Y_i | X_i = x]$$

- It turns out that this is a hard problem because we want to estimate the regression at a single point and that point is a boundary point.

General estimation strategy

- The main goal in RD is to estimate the limits of various CEFs such as:

$$\lim_{x \uparrow c} E[Y_i | X_i = x]$$

- It turns out that this is a hard problem because we want to estimate the regression at a single point and that point is a boundary point.
- As a result, the usual kinds of nonparametric estimators perform poorly (polynomials and kernels are particularly bad)

General estimation strategy

- The main goal in RD is to estimate the limits of various CEFs such as:

$$\lim_{x \uparrow c} E[Y_i | X_i = x]$$

- It turns out that this is a hard problem because we want to estimate the regression at a single point and that point is a boundary point.
- As a result, the usual kinds of nonparametric estimators perform poorly (polynomials and kernels are particularly bad)
- In general, we are going to have to choose some way of estimating the regression functions around the cutpoint.

General estimation strategy

- The main goal in RD is to estimate the limits of various CEFs such as:

$$\lim_{x \uparrow c} E[Y_i | X_i = x]$$

- It turns out that this is a hard problem because we want to estimate the regression at a single point and that point is a boundary point.
- As a result, the usual kinds of nonparametric estimators perform poorly (polynomials and kernels are particularly bad)
- In general, we are going to have to choose some way of estimating the regression functions around the cutpoint.
- Using the entire sample on either side will obviously lead to bias because those values that are far from the cutpoint are clearly different than those nearer to the cutpoint.

General estimation strategy

- The main goal in RD is to estimate the limits of various CEFs such as:

$$\lim_{x \uparrow c} E[Y_i | X_i = x]$$

- It turns out that this is a hard problem because we want to estimate the regression at a single point and that point is a boundary point.
- As a result, the usual kinds of nonparametric estimators perform poorly (polynomials and kernels are particularly bad)
- In general, we are going to have to choose some way of estimating the regression functions around the cutpoint.
- Using the entire sample on either side will obviously lead to bias because those values that are far from the cutpoint are clearly different than those nearer to the cutpoint.
- $\rightarrow$ restrict our estimation to units close to the threshold.

General estimation strategy

- The main goal in RD is to estimate the limits of various CEFs such as:

$$\lim_{x \uparrow c} E[Y_i | X_i = x]$$

- It turns out that this is a hard problem because we want to estimate the regression at a single point and that point is a boundary point.
- As a result, the usual kinds of nonparametric estimators perform poorly (polynomials and kernels are particularly bad)
- In general, we are going to have to choose some way of estimating the regression functions around the cutpoint.
- Using the entire sample on either side will obviously lead to bias because those values that are far from the cutpoint are clearly different than those nearer to the cutpoint.
- $\rightarrow$ restrict our estimation to units close to the threshold.
- Local linear regression is a good way to go: see `rdrobust` package in `R` (Calonico et al 2015)

Misconceptions

Misconceptions

- Continuity of the potential outcomes does not imply local randomization

Misconceptions

- Continuity of the potential outcomes does not imply local randomization
- This has caused a lot of confusion in the literature particularly in testing with background covariates

Misconceptions

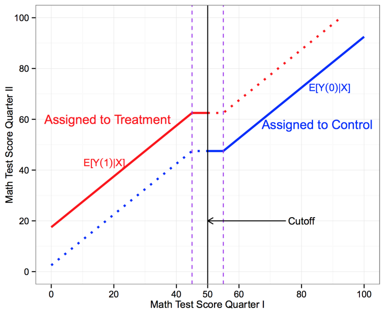
- Continuity of the potential outcomes does not imply local randomization
- This has caused a lot of confusion in the literature particularly in testing with background covariates
- Local statistical independence does not imply exclusion restriction (i.e. forcing variable not directly affecting the outcome)

Misconceptions

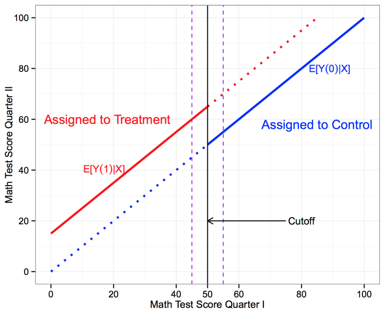
- Continuity of the potential outcomes does not imply local randomization
- This has caused a lot of confusion in the literature particularly in testing with background covariates
- Local statistical independence does not imply exclusion restriction (i.e. forcing variable not directly affecting the outcome)
- If you are doing an RDD: be sure to do balance checks and sensitivity checks (read-up on best practices first!)

Local Randomization vs. Continuity (Sekhon and Titiunik 2017)

Figure 1: Two Scenarios with Randomly Assigned Score



(a) Test scores locally unrelated to potential outcomes



(b) Test scores locally related to potential outcomes

Fuzzy RD

- With fuzzy RD, the treatment assignment is no longer a deterministic function of the forcing variable, but there is still a discontinuity in the probability of treatment at the threshold:

Assumption FRD

$$\lim_{x \downarrow c} \Pr[D_i = 1 | X_i = x] \neq \lim_{x \uparrow c} \Pr[D_i = 1 | X_i = x]$$

Fuzzy RD

- With fuzzy RD, the treatment assignment is no longer a deterministic function of the forcing variable, but there is still a discontinuity in the probability of treatment at the threshold:

Assumption FRD

$$\lim_{x \downarrow c} \Pr[D_i = 1 | X_i = x] \neq \lim_{x \uparrow c} \Pr[D_i = 1 | X_i = x]$$

- In the sharp RD, this is also true, but it further required the jump in probability to be from 0 to 1.

Fuzzy RD

- With fuzzy RD, the treatment assignment is no longer a deterministic function of the forcing variable, but there is still a discontinuity in the probability of treatment at the threshold:

Assumption FRD

$$\lim_{x \downarrow c} \Pr[D_i = 1 | X_i = x] \neq \lim_{x \uparrow c} \Pr[D_i = 1 | X_i = x]$$

- In the sharp RD, this is also true, but it further required the jump in probability to be from 0 to 1.
- Fuzzy RD is often useful when the a threshold encourages participation in program, but does not actually force units to participate.

Fuzzy RD

- With fuzzy RD, the treatment assignment is no longer a deterministic function of the forcing variable, but there is still a discontinuity in the probability of treatment at the threshold:

Assumption FRD

$$\lim_{x \downarrow c} \Pr[D_i = 1 | X_i = x] \neq \lim_{x \uparrow c} \Pr[D_i = 1 | X_i = x]$$

- In the sharp RD, this is also true, but it further required the jump in probability to be from 0 to 1.
- Fuzzy RD is often useful when the a threshold encourages participation in program, but does not actually force units to participate.
- Sound familiar? Fuzzy RD is just IV!

Fuzzy RD is IV

- Forcing variable is an instrument: affects Y_i, but only through D_i (at the threshold)

Fuzzy RD is IV

- Forcing variable is an instrument: affects Y_i, but only through D_i (at the threshold)
- Let $D_i(x)$ be the potential value of treatment when we set the forcing variable to x, for some small neighborhood around c.

Fuzzy RD is IV

- Forcing variable is an instrument: affects Y_i, but only through D_i (at the threshold)
- Let $D_i(x)$ be the potential value of treatment when we set the forcing variable to x, for some small neighborhood around c.
- $D_i(x) = 1$ if unit i would take treatment when X_i was x

Fuzzy RD is IV

- Forcing variable is an instrument: affects Y_i, but only through D_i (at the threshold)
- Let $D_i(x)$ be the potential value of treatment when we set the forcing variable to x, for some small neighborhood around c.
- $D_i(x) = 1$ if unit i would take treatment when X_i was x
- $D_i(x) = 0$ if unit i would take control when X_i was x

Fuzzy RD assumptions

Assumption 2: Monotoncity

There exists ε such that $D_i(c + e) \geq D_i(c - e)$ for all $0 < e < \varepsilon$

Fuzzy RD assumptions

Assumption 2: Monotoncity

There exists ε such that $D_i(c+e) \geq D_i(c-e)$ for all $0 < e < \varepsilon$

No one is discouraged from taking the treatment by crossing the threshold.

Fuzzy RD assumptions

Assumption 2: Monotoncity

There exists ε such that $D_i(c+e) \geq D_i(c-e)$ for all $0 < e < \varepsilon$

No one is discouraged from taking the treatment by crossing the threshold.

Assumption 3: Local Exogeneity of Forcing Variable

In a neighborhood of c,

$$\{\tau_i, D_i(x)\} \perp\!\!\!\perp X_i$$

Fuzzy RD assumptions

Assumption 2: Monotoncity

There exists ε such that $D_i(c+e) \geq D_i(c-e)$ for all $0 < e < \varepsilon$

No one is discouraged from taking the treatment by crossing the threshold.

Assumption 3: Local Exogeneity of Forcing Variable

In a neighborhood of c,

$$\{\tau_i, D_i(x)\} \perp\!\!\!\perp X_i$$

Basically, in an ε-ball around c, the forcing variable is randomly assigned.

Example: Early Release Program (HDC)

- Prison system in many countries is faced with overcrowding and high recidivism rates after release.

Example: Early Release Program (HDC)

- Prison system in many countries is faced with overcrowding and high recidivism rates after release.
- Early discharge of prisoners on electronic monitoring has become a popular policy

Example: Early Release Program (HDC)

- Prison system in many countries is faced with overcrowding and high recidivism rates after release.
- Early discharge of prisoners on electronic monitoring has become a popular policy
- Difficult to estimate impact of early release program on future criminal behavior: best behaved inmates are usually the ones to be released early

Example: Early Release Program (HDC)

- Prison system in many countries is faced with overcrowding and high recidivism rates after release.
- Early discharge of prisoners on electronic monitoring has become a popular policy
- Difficult to estimate impact of early release program on future criminal behavior: best behaved inmates are usually the ones to be released early
- Marie (2008) considers Home Detention Curfew (HDC) scheme in England and Wales:

Example: Early Release Program (HDC)

- Prison system in many countries is faced with overcrowding and high recidivism rates after release.
- Early discharge of prisoners on electronic monitoring has become a popular policy
- Difficult to estimate impact of early release program on future criminal behavior: best behaved inmates are usually the ones to be released early
- Marie (2008) considers Home Detention Curfew (HDC) scheme in England and Wales:
- Fuzzy RDD: Only offenders sentenced to more than three months (88 days) in prison are eligible for HDC, but not all those with longer sentences are offered HDC

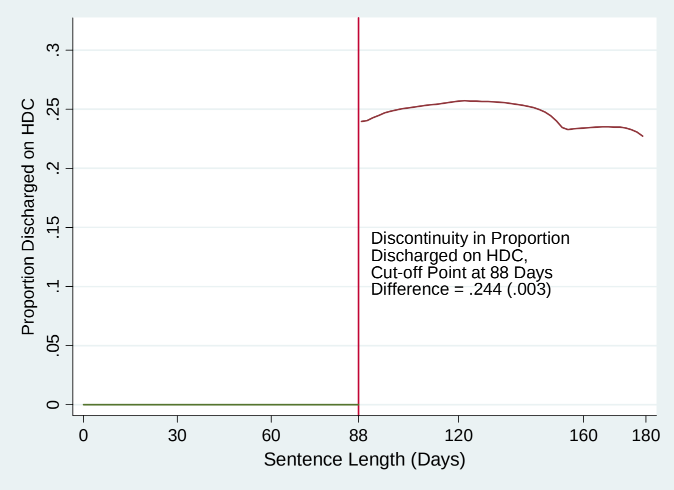
Example: Early Release Program (HDC)

Table 2: Descriptive Statistics for Prisoners Released by Length of Sentence and HDC and Non HDC Discharges and +/-7 Days Around Discontinuity Threshold

Panel A - Released +/- 7 Days of 3 Months (88 Days) Cut-off:			
Discharge Type	**Non HDC**	**HDC**	**Total**
Percentage Female	10.5	9.7	10.3
Mean Age at Release	28.9	30.7	29.3
Percentage Incarcerated for Violence	19.8	18.2	19.4
Mean Number Previous Offences	9.5	5.7	8.7
Recidivism within 12 Months	54.6	28.1	48.8
Sample Size	18,928	5,351	24,279
Panel B - Released +/- 7 Days of 3 Months (88 Days) Cu-off:			
Day of Release around Cut-off	**- 7 Days**	**+ 7 Days**	**Total**
Percentage Female	11	10.2	10.3
Mean Age at Release	28.8	29.4	29.3
Percentage Incarcerated for Violence	17.1	19.7	19.4
Mean Number Previous Offences	9.1	8.6	8.7
Recidivism within 12 Months	56.8	47.9	48.8
Percentage Released on HDC	0	24.4	22
Sample Size	2,333	21,946	24,279

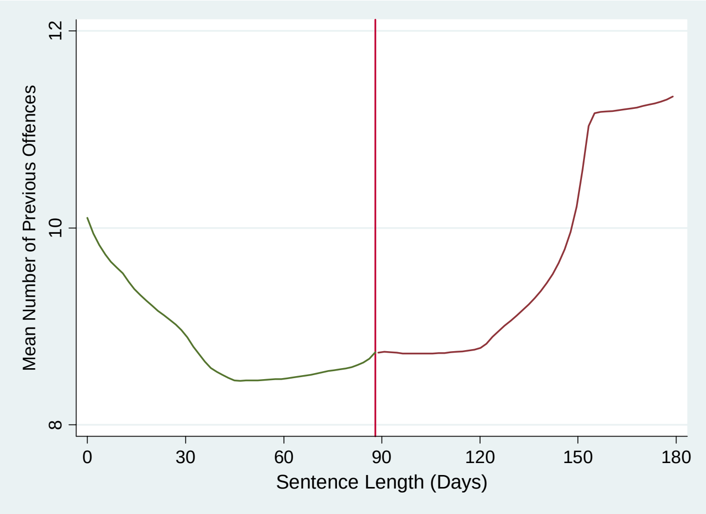
Example: Early Release Program (HDC)

Figure 1: Proportion Discharged on HDC by Sentence Length



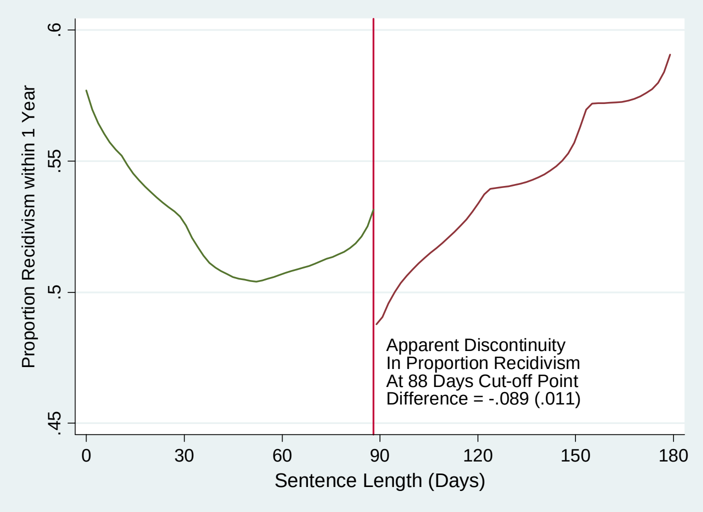
Example: Early Release Program (HDC)

Figure 2: Mean Number of Previous Offence by Sentence Length



Example: Early Release Program (HDC)

Figure 4: Recidivism within 1 Year by Sentence Length

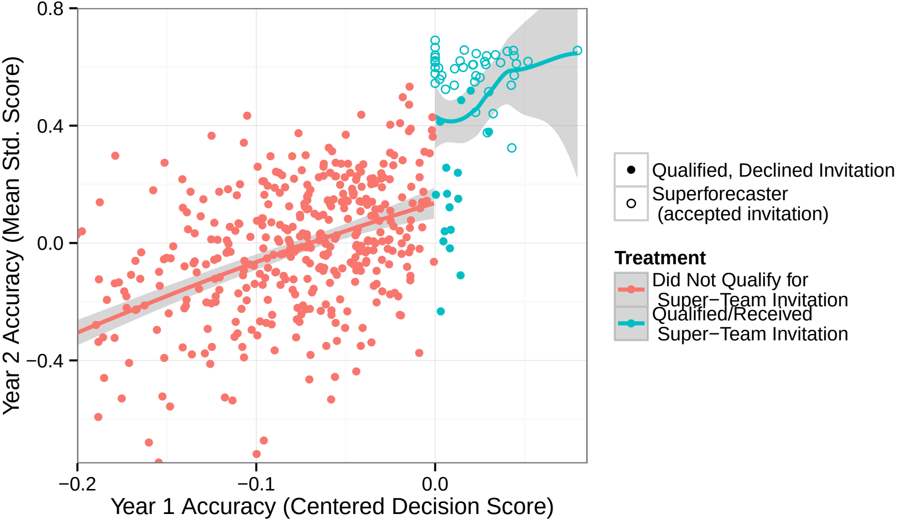


Example: Early Release Program (HDC)

Table 4: RDD Estimates of HDC Impact on Recidivism – Around Threshold

	Dependent Variable = Recidivism Within 12 Months Estimation on Individuals Discharged +/- 7 Days of 88 Days Threshold		
	(1)	(2)	(3)
Estimated Discontinuity of HDC Participation at Threshold ($HDC^{+} - HDC^{-}$)	.243 (.009)	.223 (.009)	.243 (.003)
Estimated Difference in Recidivism Around Threshold ($Rec^{+} - Rec^{-}$)	-.089 (.011)	-.059 (.009)	-.044 (.014)
Estimated Effect of HDC on Recidivism Participation ($Rec^{+} - Rec^{-}$)/ ($HDC^{+} - HDC^{-}$)	-.366 (.044)	-.268 (.044)	-.181 (n.a.)
Controls	No	Yes	No
PSM	No	No	Yes
Sample Size	24,279	24,279	24,279

Example: Teamwork



Regression Discontinuity Conclusions

- Key idea is to exploit an arbitrary assignment rule to identify a causal quantity.

Regression Discontinuity Conclusions

- Key idea is to exploit an arbitrary assignment rule to identify a causal quantity.
- Remember that we are only identifying an effect at the boundary.

Regression Discontinuity Conclusions

- Key idea is to exploit an arbitrary assignment rule to identify a causal quantity.
- Remember that we are only identifying an effect at the boundary.
- There are many other nuances to estimation and choosing an appropriate bandwidth for the comparison- be sure to read more before trying this at home.

Regression Discontinuity Conclusions

- Key idea is to exploit an arbitrary assignment rule to identify a causal quantity.
- Remember that we are only identifying an effect at the boundary.
- There are many other nuances to estimation and choosing an appropriate bandwidth for the comparison- be sure to read more before trying this at home.
- There is an interesting literature on geographic regression discontinuity designs as well. These are harder but can be useful!

Conclusion

- This week we covered approaches to unmeasured confounding

Conclusion

- This week we covered approaches to unmeasured confounding
- The trick is to exploit some other feature (No Free Lunch!)

Conclusion

- This week we covered approaches to unmeasured confounding
- The trick is to exploit some other feature (No Free Lunch!)
- Now that you have seen a few examples, hopefully you can be on the lookout for your own research.

Conclusion

- This week we covered approaches to unmeasured confounding
- The trick is to exploit some other feature (No Free Lunch!)
- Now that you have seen a few examples, hopefully you can be on the lookout for your own research.
- We talked about natural experiments, instrumental variables and regression discontinuity

Conclusion

- This week we covered approaches to unmeasured confounding
- The trick is to exploit some other feature (No Free Lunch!)
- Now that you have seen a few examples, hopefully you can be on the lookout for your own research.
- We talked about natural experiments, instrumental variables and regression discontinuity
- Next week we will talk about more designs for unmeasured confounding.

Next Week

- Causality with Repeated Data

Next Week

- Causality with Repeated Data
- Reading
  - Angrist and Pishke Chapter 5 Parallel Worlds: Fixed Effects, Differences-in-Differences and Panel Data
  - Optional: Imai and Kim "When Should We Use Linear Fixed Effects Regression Models for Causal Inference with Longitudinal Data"
  - Optional: Angrist and Pishke Chapter 6 Regression Discontinuity Designs
  - Optional: Morgan and Winship Chapter 11 Repeated Observations and the Estimation of Causal Effects

Next Week

- Causality with Repeated Data
- Reading
  - Angrist and Pishke Chapter 5 Parallel Worlds: Fixed Effects, Differences-in-Differences and Panel Data
  - Optional: Imai and Kim "When Should We Use Linear Fixed Effects Regression Models for Causal Inference with Longitudinal Data"
  - Optional: Angrist and Pishke Chapter 6 Regression Discontinuity Designs
  - Optional: Morgan and Winship Chapter 11 Repeated Observations and the Estimation of Causal Effects
- Last day of class plans

References

- Abadie, Alberto. "Semiparametric instrumental variable estimation of treatment response models." *Journal of econometrics* 113, no. 2 (2003): 231-263.
- Angrist, Joshua D., and Jrn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2008.
- Morgan, Stephen L., and Christopher Winship. *Counterfactuals and causal inference*. Cambridge University Press, 2014.

Fun with Extremists

Fun with Extremists

Hall, Andrew. "What Happens When Extremists Win Primaries?" 2015. *American Political Science Review*.

I'm grateful to Andy Hall for sharing the following slides with me.

What are the Effects of Extremists Winning Primaries?

What are the Effects of Extremists Winning Primaries?

"...getting a general-election candidate who can win is the only thing we care about."
—*Nat'l Republican Senatorial Committee*

vs.

"The road to hell is paved with electable candidates."
—*Conservative Blogger*

There is a tradeoff between ideology and electability:

There is a tradeoff between ideology and electability:

- Evaluates how the preferences of primary voters map to legislature.

There is a tradeoff between ideology and electability:

- Evaluates how the preferences of primary voters map to legislature.
- Shows how general elections react to moderates vs. extremists.

Findings: Elections Strongly Prefer Moderates

Findings: Elections Strongly Prefer Moderates

In the U.S. House, 1980–2010:

Findings: Elections Strongly Prefer Moderates

In the U.S. House, 1980–2010:

- Extremist causes 38 percentage-point decrease in win probability on average.

Findings: Elections Strongly Prefer Moderates

In the U.S. House, 1980–2010:

- Extremist causes 38 percentage-point decrease in win probability on average.
- On average, roll-call voting farther away from primary voters when they nominate extremists.

Elections Select Moderate Extremists

Elections Select Moderate Extremists

- Primary voters cannot force in extremists.

Elections Select Moderate Extremists

- Primary voters cannot force in extremists.
- House elections choose moderates, but constrained by candidate pool.

Elections Select Moderate Extremists

- Primary voters cannot force in extremists.
- House elections choose moderates, but constrained by candidate pool.
- Argument of broader research project: candidate entry key to electing extremist legislators.

Empirical Approach

Empirical Approach

- Quantity of interest: effect of extremist nominees

Empirical Approach

- Quantity of interest: effect of extremist nominees
- Ideal experiment: randomly assign districts extremist or moderate nominees.

Empirical Approach

- Quantity of interest: effect of extremist nominees

- Ideal experiment: randomly assign districts extremist or moderate nominees.

- Compare elections and roll-call voting in “treated” districts vs. “control” districts.

Obstacle to Estimating Effects of Extremist Nominees

Obstacle to Estimating Effects of Extremist Nominees

Selection Bias.

Obstacle to Estimating Effects of Extremist Nominees

Selection Bias.

- Districts choose extremist nominees because they prefer them.

Close Primaries Offer Variation in Nominee Type

Close Primaries Offer Variation in Nominee Type

- Regression discontinuity design (RDD) in primary elections.

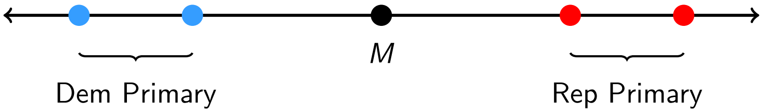
Close Primaries Offer Variation in Nominee Type

- Regression discontinuity design (RDD) in primary elections.
- Districts with moderate/extremist nominee otherwise identical in expectation.

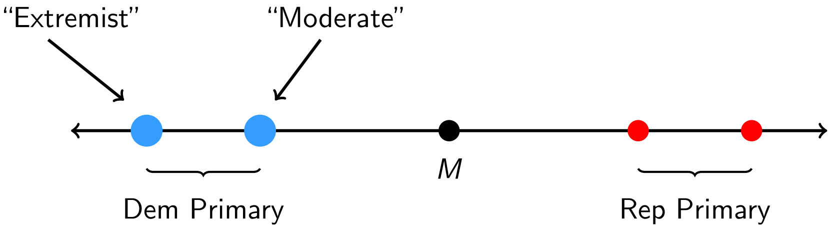
Close Primaries Offer Variation in Nominee Type

- Regression discontinuity design (RDD) in primary elections.
- Districts with moderate/extremist nominee otherwise identical in expectation.
- Key assumption for RDD: no sorting

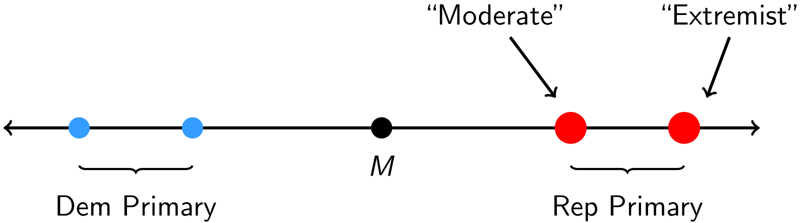
"Extremists" Defined



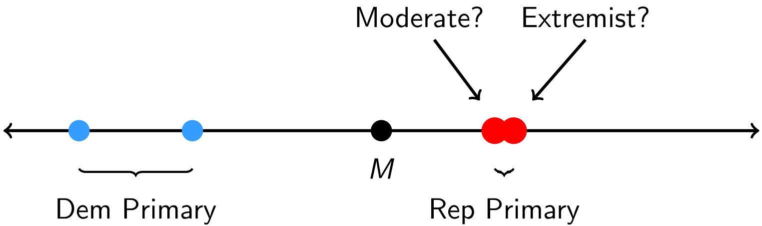
"Extremists" Defined



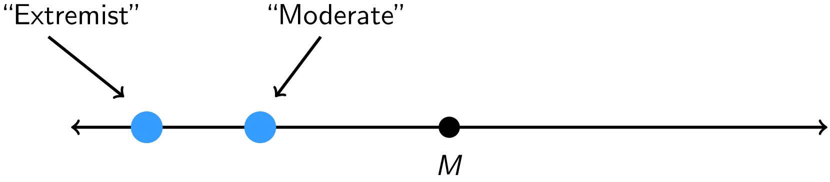
"Extremists" Defined



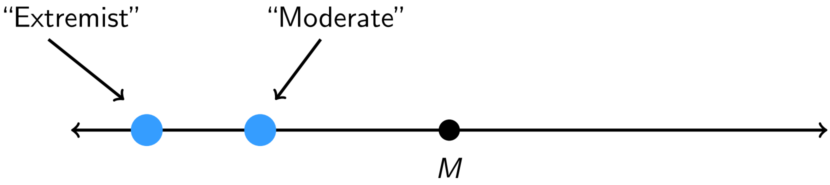
“Extremists” Defined



"Extremists" Defined

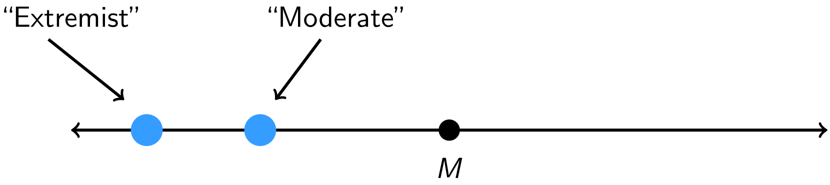


"Extremists" Defined



- Calculate distance between moderate and extremist.

"Extremists" Defined



- Calculate distance between moderate and extremist.
- Use races where distance is at or above the median distance.

Quick Example: Robbie Wills vs. Joyce Elliott

Joyce Elliott: -0.33

vs.

Robbie Wills: -0.07

Quick Example: Robbie Wills vs. Joyce Elliott

Joyce Elliott: -0.33

vs.

Robbie Wills: -0.07

Quick Example: Robbie Wills vs. Joyce Elliott

Joyce Elliott: -0.33

vs.

Robbie Wills: -0.07

- Wills sent out mailer calling Elliott an “extremist” who was “unelectable.”

Quick Example: Robbie Wills vs. Joyce Elliott

Joyce Elliott: -0.33

vs.

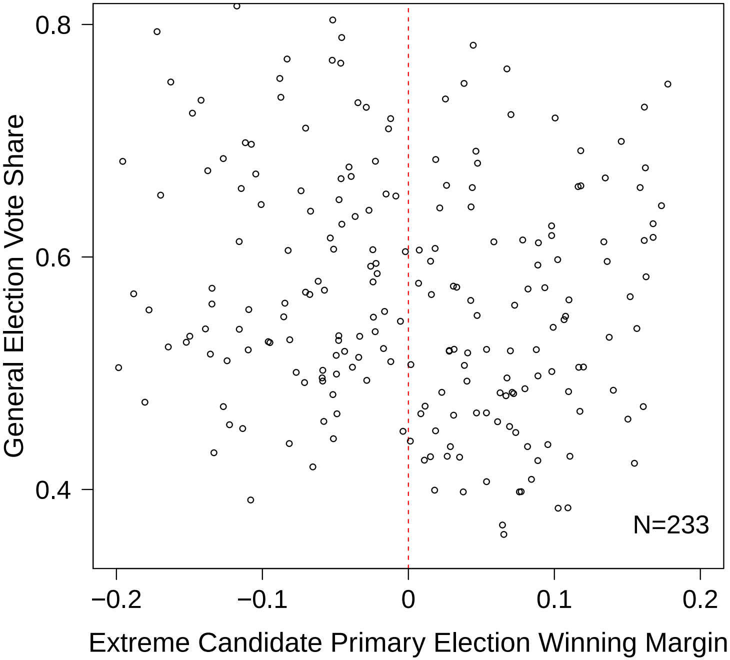
Robbie Wills: -0.07

- Wills sent out mailer calling Elliott an “extremist” who was “unelectable.”
- Elliott won close runoff primary and lost general election 62% to 38%.

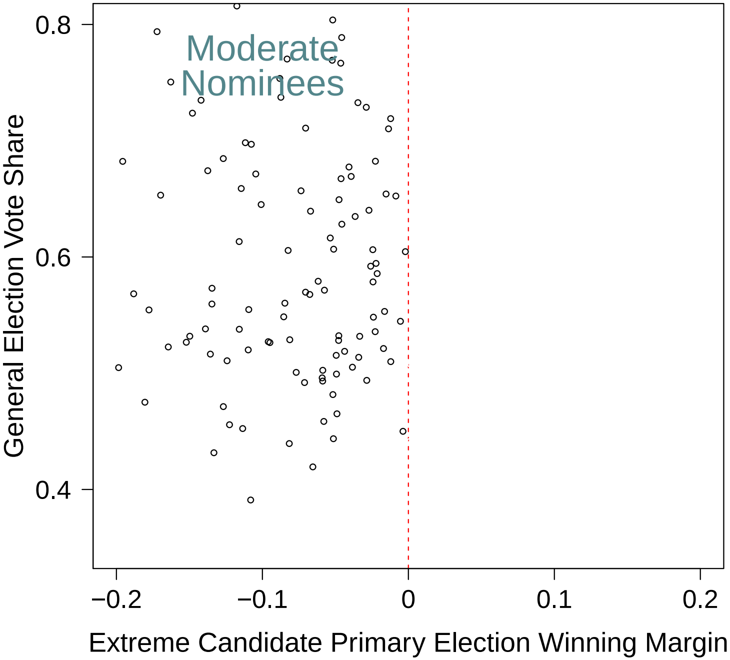
Estimating the RD: Effects of Extremist Nominations

$$Y_{it} = \beta_0 + \beta_1 Extremist\ Primary\ Win_{it} + f(V_{it}) + \epsilon_{it}$$

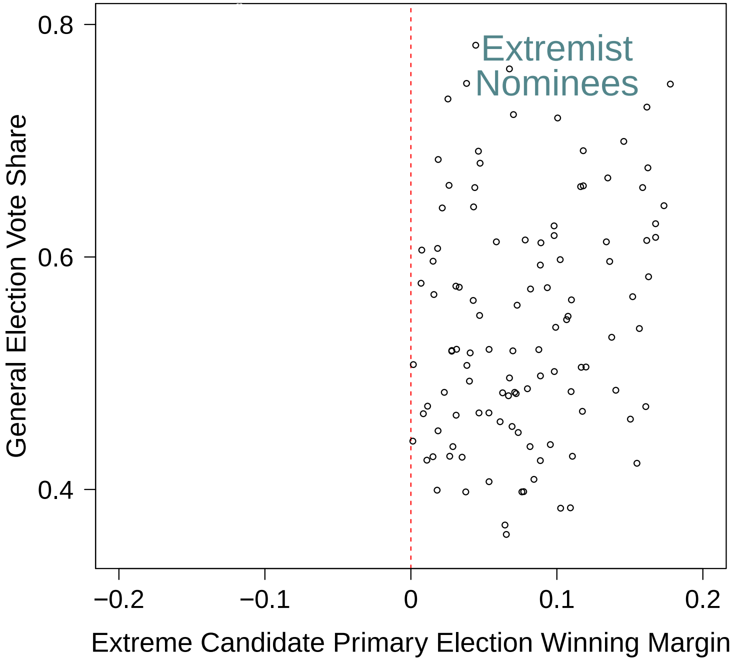
$V_{it} \equiv$ extremist candidate's vote-share winning margin.



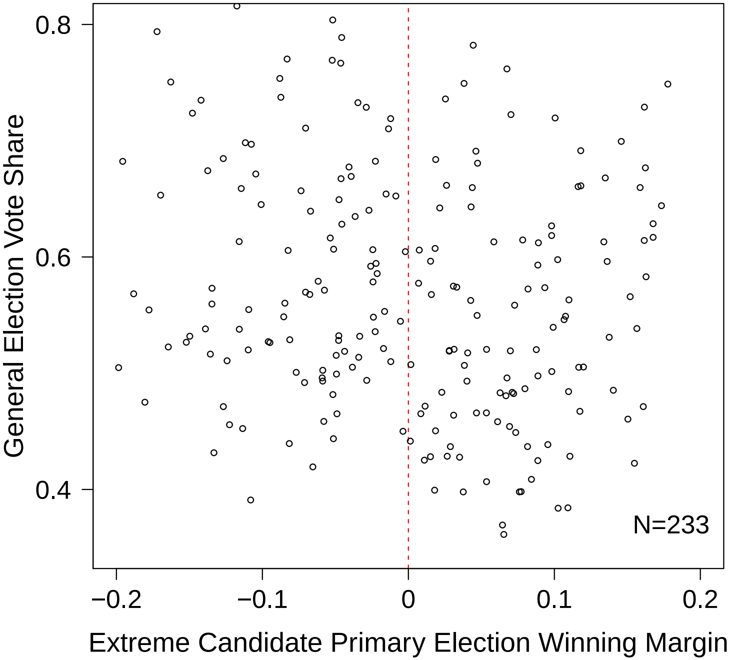
0.8
0.6
0.4
General Election Vote Share
−0.2
−0.1
0
0.1
0.2
Extreme Candidate Primary Election Winning Margin
N=233

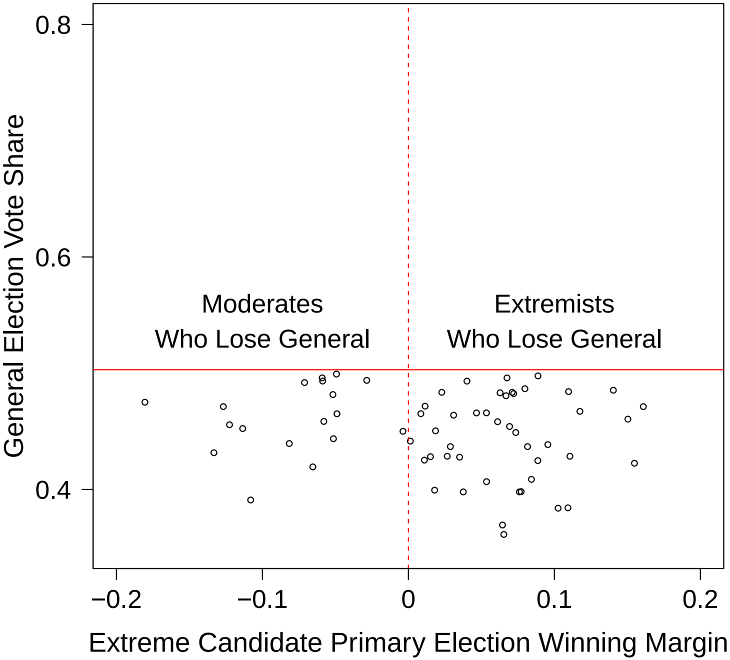


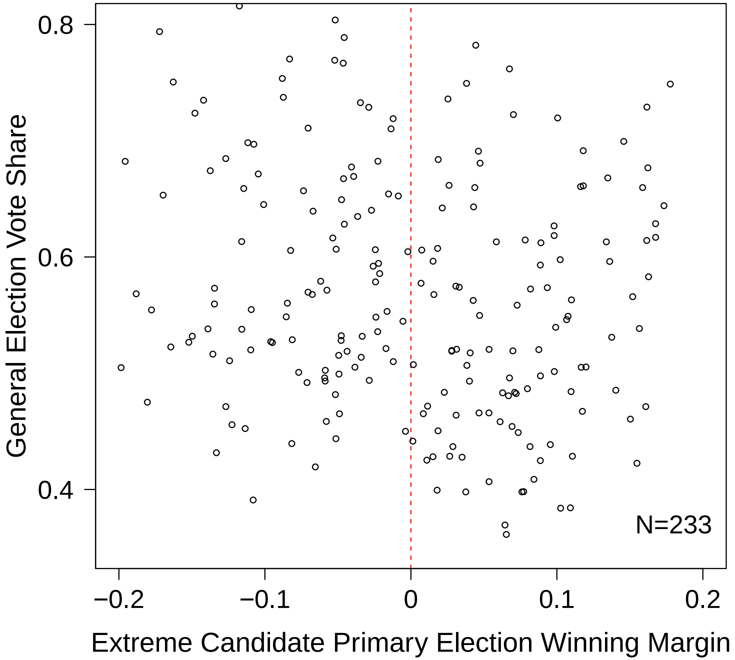
Moderate
Nominees
General Election Vote Share
0.8
0.6
0.4
−0.2
−0.1
0
0.1
0.2
Extreme Candidate Primary Election Winning Margin



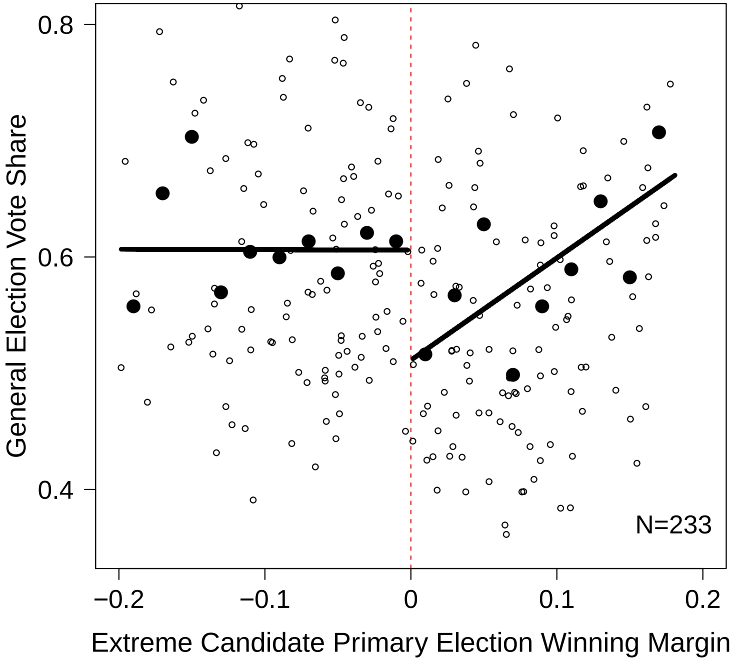
Extremist
Nominees
General Election Vote Share
0.8
0.6
0.4
−0.2
−0.1
0
0.1
0.2
Extreme Candidate Primary Election Winning Margin


0.8
0.6
0.4
General Election Vote Share
−0.2
−0.1
0
0.1
0.2
Extreme Candidate Primary Election Winning Margin
N=233

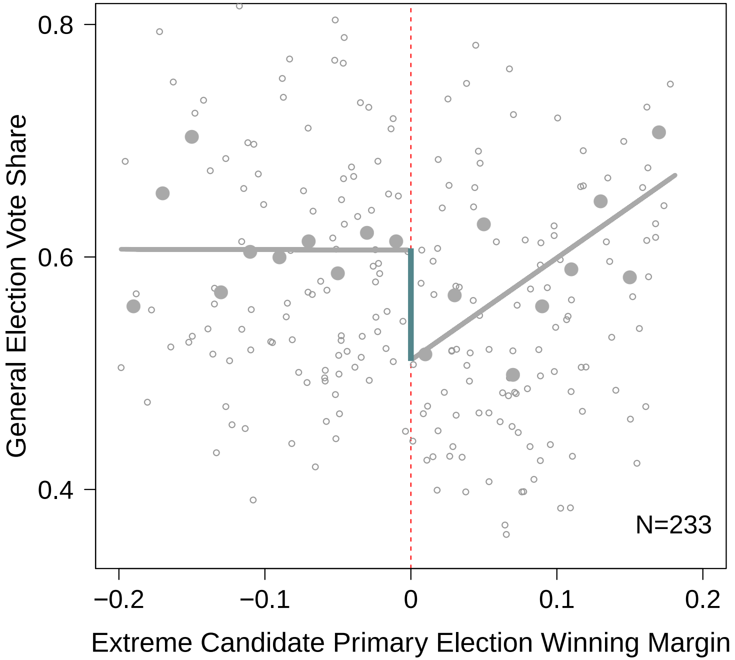

0.8
0.6
0.4
General Election Vote Share
Moderates
Who Lose General
Extremists
Who Lose General
−0.2
−0.1
0
0.1
0.2
Extreme Candidate Primary Election Winning Margin



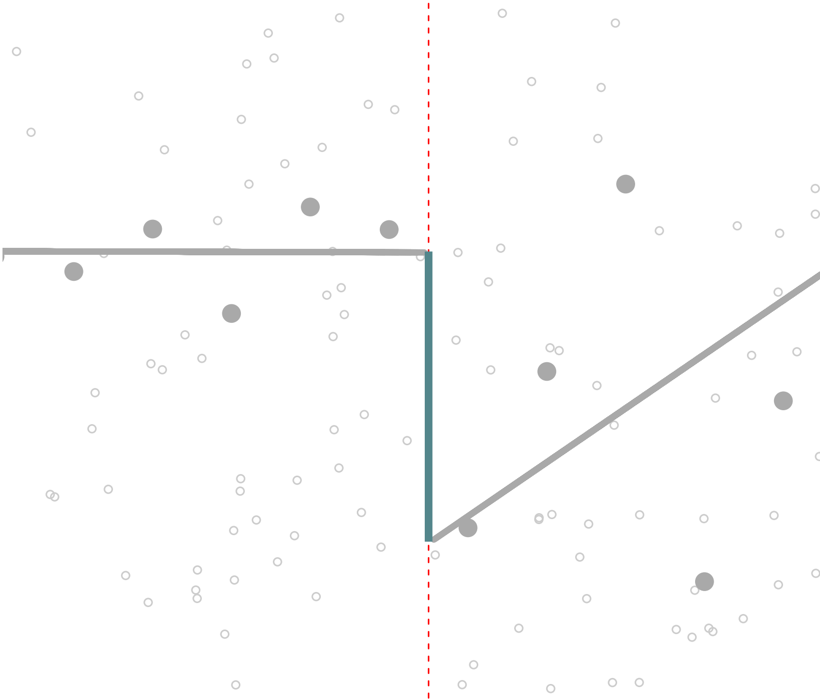
0.8
0.6
0.4
General Election Vote Share
−0.2
−0.1
0
0.1
0.2
Extreme Candidate Primary Election Winning Margin
N=233

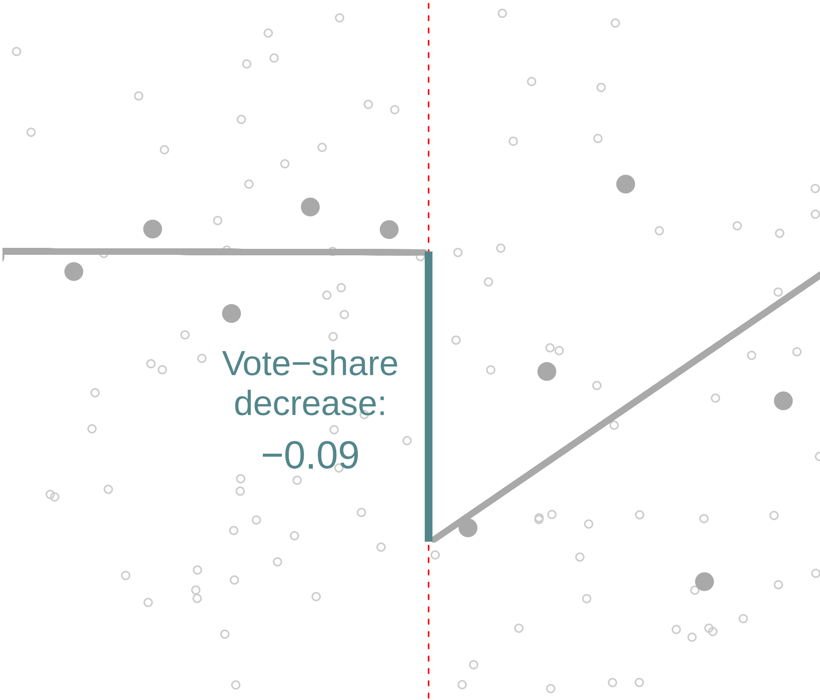


0.8
0.6
0.4
General Election Vote Share
−0.2
−0.1
0
0.1
0.2
Extreme Candidate Primary Election Winning Margin
N=233



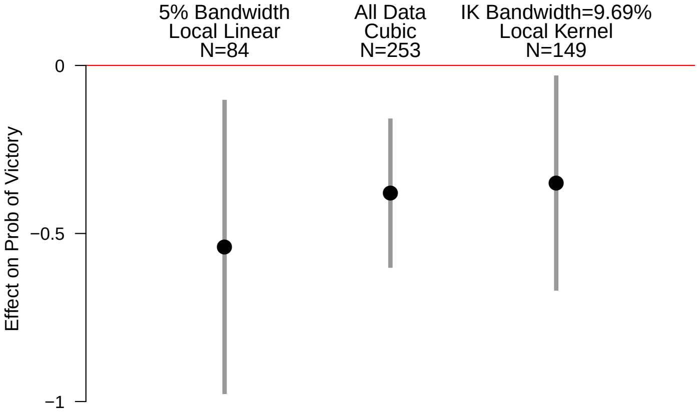
0.8
0.6
0.4
General Election Vote Share
N=233
−0.2
−0.1
0
0.1
0.2
Extreme Candidate Primary Election Winning Margin




Vote−share
decrease:
−0.09

Large Electoral Penalty to Nominating Extremist

Large Electoral Penalty to Nominating Extremist



95% Confidence Intervals From Max of Robust and Conventional Standard Errors

How Does Penalty to Extremists Affect Roll-Calls?

How Does Penalty to Extremists Affect Roll-Calls?

1. Penalty makes other party more likely to win seat.

How Does Penalty to Extremists Affect Roll-Calls?

1. Penalty makes other party more likely to win seat.
2. Extremist offers more extreme roll-call voting.

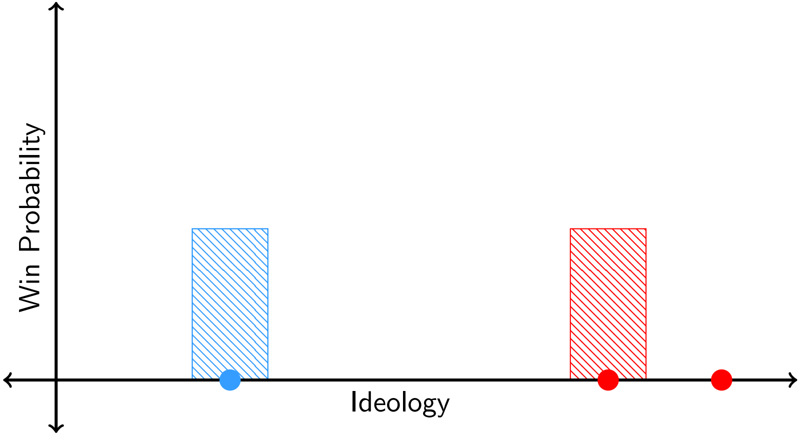
How Does Penalty to Extremists Affect Roll-Calls?

1. Penalty makes other party more likely to win seat.
2. Extremist offers more extreme roll-call voting.

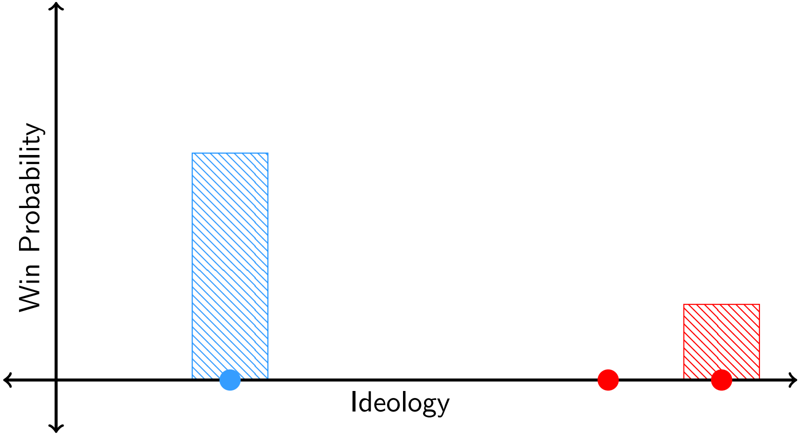
Knowing general election prefers moderates not sufficient to understand tradeoff.

How Does Penalty to Extremists Affect Roll-Calls?

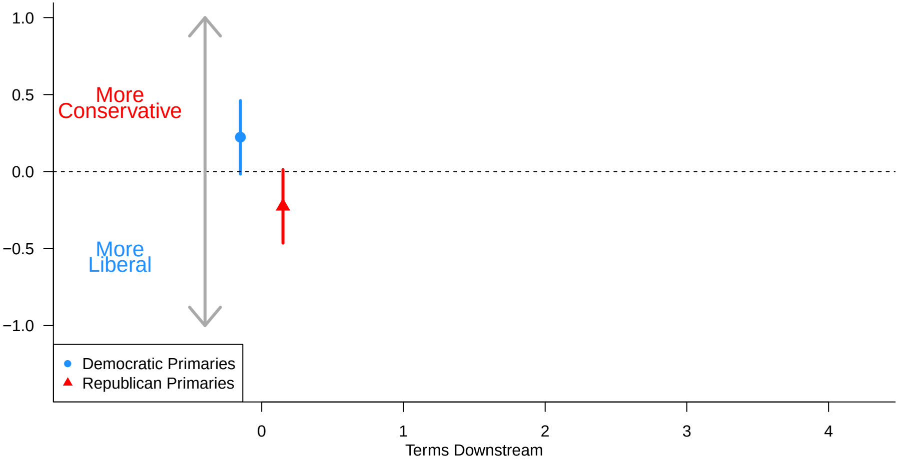
How Does Penalty to Extremists Affect Roll-Calls?



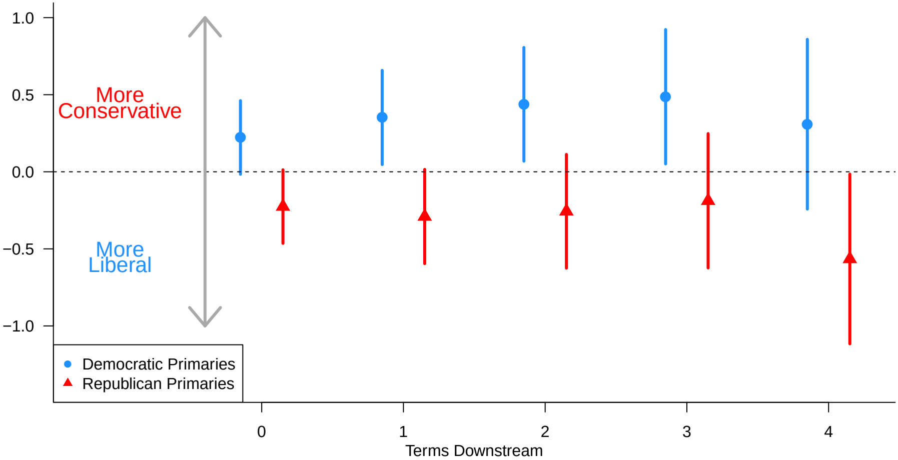
How Does Penalty to Extremists Affect Roll-Calls?



Effect of Extremists on Roll-Call Voting



Effect of Extremists on Roll-Call Voting



Summary

Summary

- Primary voters do not make legislature more extreme by forcing in extreme candidates.

Summary

- Primary voters do not make legislature more extreme by forcing in extreme candidates.
- The general election is a huge force for moderation.

Elections: A Limited Force For Moderation

Elections: A Limited Force For Moderation

- U.S. House elections select "moderate extremists."

Elections: A Limited Force For Moderation

- U.S. House elections select “moderate extremists.”

- Argument: Differential entry of extremist candidates forces voters to elect extremists.

Fun With Related Work

Hall and Snyder. 2013. Candidate Ideology and Electoral Success. Working Paper.

Eggers, Andrew, Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder, Jr. On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: Evidence From Over 40,000 Close Races. *American Journal of Political Science*, 2015.

Hall, Andrew B. "What Happens When Extremists Win Primaries?" *American Political Science Review*. 2015.

Proof of the LATE theorem

- Under the exclusion restriction and randomization,

$$\begin{aligned} E[Y_i|Z_i = 1] &= E[Y_i(0) + (Y_i(1) - Y_i(0))D_i|Z_i = 1] \\ &= E[Y_i(0) + (Y_i(1) - Y_i(0))D_i(1)] \qquad \text{(randomization)} \end{aligned}$$

- The same applies to when $Z_i = 0$, so we have

$$E[Y_i|Z_i = 0] = E[Y_i(0) + (Y_i(1) - Y_i(0))D_i(0)]$$

- Thus, $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] =$

$$\begin{aligned} &E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] \\ =&E[(Y_i(1) - Y_i(0))(1)|D_i(1) > D_i(0)] \Pr[D_i(1) > D_i(0)] \\ +&E[(Y_i(1) - Y_i(0))(-1)|D_i(1) < D_i(0)] \Pr[D_i(1) < D_i(0)] \\ =&E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)] \Pr[D_i(1) > D_i(0)] \end{aligned}$$

- The third equality comes from monotonicity: with this assumption, $D_i(1) < D_i(0)$ never occurs.

Proof (continued)

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)] \Pr[D_i(1) > D_i(0)]$$

- We can use the same argument for the denominator:

$$\begin{aligned} E[D_i|Z_i = 1] - E[D_i|Z_i = 0] &= E[D_i(1) - D_i(0)] \\ &= \Pr[D_i(1) > D_i(0)] \end{aligned}$$

- Dividing these two expressions through gives the LATE.