Week 11: Causality with Unmeasured Confounding

Brandon Stewart

Princeton

December 5 and 7, 2016

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

---
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?
Approaches to Unmeasured Confounding

Natural Experiments

Traditional Instrumental Variables

Fun with Coarsening Bias

Modern Approaches to Instrumental Variables

Regression Discontinuity

Fun with Extremists

Fun With Weak Instruments

Appendix
Unmeasured Confounding

- Last week we considered cases of measured confounding
  \[ X \downarrow \downarrow \]
  \[ D \rightarrow Y \]
  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.
- No Free Lunch \( \Leftarrow \) 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 more research if you want to use one of these techniques.
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
- Synthetic Controls
1 Approaches to Unmeasured Confounding

2 Natural Experiments

3 Traditional Instrumental Variables

4 Fun with Coarsening Bias

5 Modern Approaches to Instrumental Variables

6 Regression Discontinuity

7 Fun with Extremists

8 Fun With Weak Instruments

9 Appendix
Natural Experiments

- Broadly speaking an “experiment” where the treatment is randomized but the randomization was not controlled by the researcher.
- 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, an excellent way to capitalize on randomness in the world to make casual inferences.
- See Dunning (2012) *Natural Experiments in the Social Sciences*
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)

<table>
<thead>
<tr>
<th>Randomness</th>
<th>Focus</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vietnam draft</td>
<td>labor market</td>
<td>Angrist 1990</td>
</tr>
<tr>
<td>randomized quotas</td>
<td>female leadership in Indian village council presidencies</td>
<td>Chattopadhyay &amp; Duflo 2004</td>
</tr>
<tr>
<td>randomized term lengths</td>
<td>tenure in office on legislative performance</td>
<td>Dal Bo &amp; Rossi 2010</td>
</tr>
<tr>
<td>lottery</td>
<td>effect of winnings on political attitudes</td>
<td>Doherty, Green &amp; Gerber 2006</td>
</tr>
<tr>
<td>randomized ballot order</td>
<td>ballot order effects in CA</td>
<td>Ho &amp; Imai 2008</td>
</tr>
</tbody>
</table>
### Natural Experiment Examples (As If Randomization)

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

From Sekhon and Titiunik (2012)

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

FIGURE 1. Illustration of Redistricting Research Designs

(a) Before One-time Redistricting  
(b) After One-time Redistricting
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 I want?)
- The result only applies to a limited population that we may not care a great deal about.
- Be sure to verify that you “as-if-random” assignment is really random (e.g. placebo tests, balance tests)
- Convincingly analyzing a natural experiment takes at least as much careful thought not less!
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

- 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.
Approaches to Unmeasured Confounding

Natural Experiments

Traditional Instrumental Variables

Fun with Coarsening Bias

Modern Approaches to Instrumental Variables

Regression Discontinuity

Fun with Extremists

Fun With Weak Instruments

Appendix
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.
Graphical Model

- $Z$ is the instrument, $D$ is the treatment, and $U$ is the unmeasured confounder.
- Exclusion restriction
  - no common causes of the instrument and the outcome
  - no direct or indirect effect of the instrument on the outcome not through the treatment
- First-stage relationship $Z$ affects $D$
Some Examples

- Angrist (1990): Draft lottery as an IV for military service (income 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)
- Acharya, Blackwell, Sen (2015): cotton suitability as IV for proportion slave in 1860 (outcome is white attitudes today)
Core Idea

The world has randomized something just maybe not the thing you want.

Subject to an exclusion restriction you may be able to get (approximately) what you want anyway.
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

<table>
<thead>
<tr>
<th></th>
<th>Not Enrolled in Training</th>
<th>Enrolled in Training</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assigned to Control</td>
<td>3,663</td>
<td>54</td>
<td>3,717</td>
</tr>
<tr>
<td>Assigned to Training</td>
<td>2,683</td>
<td>4,804</td>
<td>7,487</td>
</tr>
<tr>
<td>Total</td>
<td>6,346</td>
<td>4,858</td>
<td>11,204</td>
</tr>
</tbody>
</table>
Two Views on Instrumental Variables

1. Traditional Econometric Framework
   - Constant treatment effects
   - Linearity in case of a multivalued treatment

2. Potential Outcome Model of IV
   - Heterogeneous treatment effects
   - Focus in Local Average Treatment Effect (LATE)
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)

Suppose that the classical “exogeneity” condition \( E[U|X_1, ..., X_k] = 0 \) does not hold for the causally interpreted linear model:

\[
Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_k X_{ik} + U_i \\
U_i \sim_{i.i.d} N(0, \sigma_u^2).
\]

This can happen for a number of reasons including:

- omitted variables
- measurement error in X
- included variables (post-treatment or M-structures)
- simultaneous equations (endogenous feedback loops)
A Potential Solution: Instrumental Variables (IV)

\[ Y_i = \beta_0 + \beta_1 X_i + U_i \]
\[ E[U_i|X_i] \neq 0 \]
\[ X_i = \gamma_0 + \gamma_1 Z_i + U_i \]
\[ E[U_i|Z_i] = 0 \]
\[ \text{Cov}[X_i, Z_i] \neq 0 \]
Commonly Used Instrumental Variables

- Assigned status in randomized trials with noncompliance
  - the assigned versus enrolled distinction you explored in the Lalonde data
  - “received” versus “read” the mailing in the social pressure experiment
- Rainfall, earthquakes, ...
- ...

Diagram:

- $U \rightarrow Y$
- $X \rightarrow Y$
- $Z \rightarrow X$
- $X \rightarrow Y$
- $\beta_1$ and $\gamma_1$
The IV Estimator

With our assumed model,

- regressing $X$ on $Z$ identifies $\gamma_1$
- regressing $Y$ on $Z$ identifies $\gamma_1 \cdot \beta_1 = \hat{\gamma}_1 \cdot \hat{\beta}_1$
- $\frac{\gamma_1 \cdot \beta_1}{\hat{\gamma}_1}$ identifies $\frac{\gamma_1 \cdot \beta_1}{\gamma_1} = \beta_1$
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 $\gamma_1$ and $\gamma_1 \cdot \beta_1$ are not perfect, we can get inaccurate estimates of $\beta_1$:

- medium sample size $\Rightarrow$ high variance
- small violations of assumptions $\Rightarrow$ large bias
Suppose we believe that the effects of $Z$ and $X$ are different for different units.

\[ Y_i = \beta_0 + \beta_1 X_i + U_i \]

\[ X_i = \gamma_0 + \gamma_1 Z_i + V_i \]
IV Estimator with Heterogeneous Effects

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

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

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

- Recall that the OLS estimator for $\alpha_1$ is given by:

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

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

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

$$
E[\hat{\alpha}_{1,\text{OLS}}] = \alpha_1 + E\left[\frac{\text{Cov}[D, u_2]}{\text{Cov}[D, D]}\right]
$$

so bias depends on correlation between $u$ and $D$
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)
  - \( \pi_1 \) 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 \( \text{Cov}[u_1, Z] = 0 \) and \( \text{Cov}[u_2, Z] = 0 \).
   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

- Example: JTPA
  - Offer to get Training Z
  - Earnings Y
  - Training D

Stewart (Princeton)
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**: $\text{Cov}[u_1, Z] = 0$, $\pi_1 \neq 0$, and $\text{Cov}[u_2, Z] = 0$

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

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

First stage effect: \( Z \) on \( D \)

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

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

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

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

\( \hat{\pi}_1 \) is consistent since \( \text{Cov}[u_1, Z] = 0 \)
First stage effect: $Z$ on $D$: $\hat{\pi}_1 = \frac{\text{Cov}[D,Z]}{\text{V}[Z]}$

R Code

```r
> 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

```r
> 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-16
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: $\text{Cov}[u_1, Z] = 0$, $\pi_1 \neq 0$, and $\text{Cov}[u_2, Z] = 0$

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

\[ Y = \alpha_0 + \alpha_1 (\pi_0 + \pi_1 Z + u_1) + u_2 \]
\[ Y = (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1) Z + (\alpha_1 u_1 + u_2) \]
\[ Y = \gamma_0 + \gamma_1 Z + u_3 \]

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

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

$\hat{\gamma}_1$ is consistent since $\text{Cov}[u_1, Z] = 0$ and $\text{Cov}[u_2, Z] = 0$ implies $\text{Cov}[u_3, Z] = 0$.
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
-16200.000  -13802.836  -4816.957   8950.139  139560.000

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)      15040.500    274.929  54.716  < 2e-16 ***
assignmt            1159.424    336.303   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

- **Second Stage:** \( Y = \alpha_0 + \alpha_1 D + u_2 \)
- **First Stage:** \( D = \pi_0 + \pi_1 Z + u_1 \)
- **IV assumptions:** \( \text{Cov}[u_1, Z] = 0, \pi_1 \neq 0, \) and \( \text{Cov}[u_2, Z] = 0 \)

**IV Effect:** \( X \) on \( Y \) using exogenous variation in \( D \) that is induced by \( Z \). Recall

\[
Y = (\alpha_0 + \alpha_1 \pi_0) + (\alpha_1 \pi_1)Z + (\alpha_1 u_1 + u_2) \\
Y = \gamma_0 + \gamma_1 Z + u_3
\]

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 \( \alpha_1: \)

\[
\alpha_1 = \frac{\gamma_1}{\pi_1} = \frac{\text{Effect of } Z \text{ on } Y}{\text{Effect of } Z \text{ on } D} = \frac{\text{Cov}[Y, Z]/\text{Cov}[Z, Z]}{\text{Cov}[D, Z]/\text{Cov}[Z, Z]} = \frac{\text{Cov}[Y, Z]}{\text{Cov}[D, Z]} \\
\hat{\alpha}_1 = \frac{\text{Cov}[\alpha_0 + \alpha_1 D + u_2, Z]}{\text{Cov}[D, Z]} = \frac{\alpha_1 \text{Cov}[D, Z] + \text{Cov}[u_2, Z]}{\text{Cov}[D, Z]} = \alpha_1 + \frac{\text{Cov}[u_2, Z]}{\text{Cov}[D, Z]} \\
E[\hat{\alpha}_1] = \alpha_1 + E[\frac{\text{Cov}[u_2, Z]}{\text{Cov}[D, Z]}]
\]

\( \hat{\alpha}_1 \) is consistent if \( \text{Cov}[u_2, Z] = 0 \) but has a bias which decreases with instrument strength.
Instrumental Variable Effect: Wald Estimator

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

R Code

```r
> 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

```r
> 257.0625061/0.1390407
[1] 1848.829
```
Instrumental Variable Effect: Two Stage Least Squares

The instrumental variable estimator:

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

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:

\[
Y = \alpha_0 + \alpha_1 \hat{D} + u_2
\]

\[
Y = \alpha_0 + \alpha_1 (\hat{\pi}_0 + \hat{\pi}_1 Z) + u_2
\]

\[
Y = (\alpha_0 + \alpha_1 \hat{\pi}_0) + \alpha_1 (\hat{\pi}_1 Z) + u_2
\]

- \( \alpha_1 \) 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

```r
> 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

```r
> 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

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

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

so to obtain consistent estimates the instrument \( Z \) must be:

- **Relevant**: \( \text{Cov}(Z, D) \neq 0 \) (testable)
  - If \( \text{Cov}(Z, D) \) is small, the instrument is weak. We get consistency in asymptotia, but in small (finite) samples we can get strong bias even if instrument is perfectly exogenous.

- **Exogenous**: \( \text{Cov}(Z, u_2) = 0 \) (untestable)
  - If \( Z \) has an independent effect on \( Y \) other than through \( D \) we have \( \text{Cov}(Z, u_2) \neq 0 \) and estimates are inconsistent.
  - Even 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.
### Instrumental Variable Examples

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

- In contrast to OLS, the IV estimator is not unbiased in small (finite) samples even when instrument is perfectly exogenous.
- Because of sampling variability in first stage estimation of fitted values, some part of the correlation between errors in first and second stage seeps into 2SLS estimates (correlation disappears in large samples).
- Finite sample bias can be considerable (e.g., 20 - 30%), even when the sample size is over 100,000 if the instrument is weak.
- Relative bias of $\alpha_{D,IV}$ versus $\alpha_{D,OLS}$ is approximately $1/F$ where $F$ is the $F$-statistic for testing $H_0: \pi_Z = 0$, i.e. partial effect of $Z$ on $D$ is zero (or against joint zero for multiple instruments).
Exogenous, but Weak Instruments

- Adding instruments increases the relevance of the instrument set (increases the first stage $F$)
- But too many instruments increases small sample bias (compared to few instruments) and also call into doubt the exclusion restrictions
- Best to have single, strong instrument
- There are more complex competitors to 2SLS:
  - Limited Information Maximum Likelihood (LIML) estimation
  - Jackknife instrumental variables
  - Imbens and Rosenbaum (2005) robust IV.
- Small sample studies suggest that LIML and robust IV may be superior to 2SLS in small samples (but remains open area of research)
Failure of Exogeneity

- Note the probability limit:

\[
\text{plim} \alpha_{D, IV} = \alpha_D + \frac{\text{Corr}(Z, u_2) \sigma_{u_2}}{\text{Corr}(Z, D) \sigma_D}
\]

- In general we get inconsistent estimates if \( \text{Corr}(Z, u_2) \neq 0 \). This large sample bias can often be considerable but is hard to quantify precisely because it depends on unobservables.

- If the instrument is stronger, large sample bias can be attenuated, but often magnitude of \( \text{Corr}(Z, u_2) \) dominates.

- The best we can often do is often to be skeptical and to make sure exogeneity is highly plausible in the setting to which we apply IV.

- **Sensitivity analysis:**
  - Is the instrument plausibly exogenous or can it be easily predicted from covariates?
Failure of Exogeneity

- 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 designing an encouragement experiment we need to be careful to design encouragements so that they are relevant, but also narrowly targeted to only create variation in treatment intake
- SUTVA may be a concern as well
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”
- Hernan and Robins (2006)
Conclusion

- 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 exogenous

- Violations of assumptions can lead to large biases and estimation theory is complicated

- So far, we have assumed constant treatment effects $\alpha_D$ 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
Week 11: Unmeasured Confounding

1. Approaches to Unmeasured Confounding
2. Natural Experiments
3. Traditional Instrumental Variables
4. Fun with Coarsening Bias
5. Modern Approaches to Instrumental Variables
6. Regression Discontinuity
7. Fun with Extremists
8. Fun With Weak Instruments
9. Appendix
Coarsening Bias: How Coarse Treatment Measurement Upwardly Biases Instrumental Variable Estimates

John Marshall

Department of Government, Harvard University, Cambridge, MA 02138
e-mail: jlmash@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...

- **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?
1. Approaches to Unmeasured Confounding
2. Natural Experiments
3. Traditional Instrumental Variables
4. Fun with Coarsening Bias
5. Modern Approaches to Instrumental Variables
6. Regression Discontinuity
7. Fun with Extremists
8. Fun With Weak Instruments
9. Appendix
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 $\alpha = Y_{s,i} - Y_{s-1,i}$. 
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 $Z_i \leadsto$ 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

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 $Z = z$.

- $D_1 = 1$ encouraged to take treatment and takes treatment

Assumption

*Observed treatments are realized as*

$$D = Z \cdot D_1 + (1 - Z) \cdot D_0 \text{ so } D_i = \begin{cases} 
D_{1i} & \text{if } Z_i = 1 \\
D_{0i} & \text{if } Z_i = 0 
\end{cases}$$
Key Assumptions in the Modern Approach

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 need 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 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.

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

- Implies that

\[ \text{Cov}(D_i, Z_i) \neq 0, \]
\[ 0 < P(Z = 1) < 1 \]
\[ P(D_1 = 1) \neq P(D_0 = 1) \]

- This is testable by regressing \( D \) on \( Z \)

- Note that even a weak instrument can induce a lot of bias. 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.
- **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

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

**Definition**

- **Compliers**: $D_1 > D_0$ ($D_0 = 0$ and $D_1 = 1$).
- **Always-takers**: $D_1 = D_0 = 1$.
- **Never-takers**: $D_1 = D_0 = 0$.
- **Defiers**: $D_1 < D_0$ ($D_0 = 1$ and $D_1 = 0$).

Only one of the potential treatment indicators ($D_0, D_1$) is observed, so in the general case we cannot identify which group any particular individual belongs to.
Monotonicity means no defiers

<table>
<thead>
<tr>
<th>Name</th>
<th>$D_i(1)$</th>
<th>$D_i(0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always Takers</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Never Takers</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Compliers</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Defiers</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

- 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) or the complier average treatment effect (CATE).
- 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.
Who are the Compliers?

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

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

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 $\implies$ no “always-takers” and since there are no defiers,

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

Proof.

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

(exclusion restriction + one-sided noncompliance)

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

(randomization)

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

(law of iterated expectations + binary treatment)

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

(one-sided noncompliance)

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))

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

\[
P(D_1 > D_0) = E[D_1 - D_0] = E[D_1] - E[D_0]
= E[D|Z = 1] - E[D|Z = 0]
\]

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

\[
P(D_1 > D_0|D = 1) = \frac{P(Z = 1)(E[D|Z = 1] - E[D|Z = 0])}{P(D = 1)}
\]

- 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).
## Table 4.4.2
Probabilities of compliance in instrumental variables studies

<table>
<thead>
<tr>
<th>Source</th>
<th>Endogenous Variable (D)</th>
<th>Instrument (z)</th>
<th>Sample</th>
<th>( P[D = 1] ) ( (5) )</th>
<th>First Stage, ( P[D_1 &gt; D_0] ) ( (6) )</th>
<th>( P[z = 1] ) ( (7) )</th>
<th>Compliance Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Non-white men born in 1950</td>
<td>.163</td>
<td>.060</td>
<td>.534</td>
<td>.197</td>
</tr>
<tr>
<td>Angrist and Krueger (1991)</td>
<td>High school graduate</td>
<td>First two children are same sex</td>
<td>Men born between 1930 and 1939</td>
<td>.381</td>
<td>.060</td>
<td>.506</td>
<td>.080</td>
</tr>
<tr>
<td>Acemoglu and Angrist (2000)</td>
<td>High school graduate</td>
<td>Third- or fourth-quarter birth</td>
<td>State requires 11 or more years of school attendance</td>
<td>.770</td>
<td>.016</td>
<td>.509</td>
<td>.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>White men aged 40–49</td>
<td>.617</td>
<td>.037</td>
<td>.300</td>
<td>.018</td>
</tr>
</tbody>
</table>

**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.
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.
Table 7—Reduced Form Relationship between the Distance from the Coast and Trust within Africa and Asia

<table>
<thead>
<tr>
<th></th>
<th>Trust of local government council</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Afrobarometer sample</td>
<td>Asiabaraomer sample</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Distance from the coast</td>
<td>0.00039***</td>
<td>0.00031***</td>
<td>-0.00001</td>
</tr>
<tr>
<td></td>
<td>(0.00009)</td>
<td>(0.00008)</td>
<td>(0.00010)</td>
</tr>
<tr>
<td>Country fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Individual controls</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Number of observations</td>
<td>19,913</td>
<td>19,913</td>
<td>5,409</td>
</tr>
<tr>
<td>Number of clusters</td>
<td>185</td>
<td>185</td>
<td>62</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.16</td>
<td>0.18</td>
<td>0.19</td>
</tr>
</tbody>
</table>

*Notes:* The table reports OLS estimates. The unit of observation is an individual. The dependent variable in the Asiabaraomer 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 Asiabaraomer as in the Afrobarometer. Standard errors are clustered at the ethnicity level in the Afrobarometer regressions and at the location (city) level in the Asiabaraomer 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
- Covariates and conditional ignorability
- Overidentification tests (in constant effects)
- Compliance modeling for weak instruments (see Fun With Weak Instruments)
- Conditional effects and causal interaction models
We dropped the constant effects assumption, 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.)
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?  
   (does $\text{Cov}[\nu_1, Z] = 0$)
3. does the exclusion restriction hold?  
   (is there a path from $Z$ to $Y$ not through $D$)
4. do we believe monotonicity?
5. do the assumptions identify an effect for the subpopulation of interest?

- Be sure to evaluate all conditions and remember randomization of $Z$ does not guarantee the exclusion restriction.
1. Approaches to Unmeasured Confounding
2. Natural Experiments
3. Traditional Instrumental Variables
4. Fun with Coarsening Bias
5. Modern Approaches to Instrumental Variables
6. Regression Discontinuity
7. Fun with Extremists
8. Fun With Weak Instruments
9. Appendix
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).
Setup

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

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

- $X_i$ 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 $\sim$ discontinuity created by the treatment to estimate the effect of $D$ on $Y$ at the threshold
Graphical Illustration

\[ X_i = c \]
Graphical Illustration

$X_i = c$

$Y_i$

$X_i$

$Y_i$
Graphical Illustration

\[X_i = c\]

\[\text{LATE} = E[Y_1 - Y_0 | X = c]\]

\[E[Y_1 | X, D=1]\]
\[E[Y_0 | X, D=1]\]
\[E[Y_1 | X, D=0]\]
\[E[Y_0 | X, D=0]\]
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:

$$
P(D_i = 1|X_i = c) = 1
$$

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

- 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. $Y_1, Y_0 \perp D \mid X$ (trivially met)
2. $0 < P(D = 1 \mid X = x) < 1$ (always violated in Sharp RDD)
3. $E[Y_1 \mid X, D]$ and $E[Y_0 \mid X, D]$ 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:

$$\alpha_{SRDD} = E[Y_1 - Y_0 \mid X = c] = E[Y_1 \mid X = c] - E[Y_0 \mid X = c] = \lim_{x \downarrow c} E[Y_1 \mid X = x] - \lim_{x \uparrow c} E[Y_0 \mid X = x]$$

Without further assumptions $\alpha_{SRDD}$ is only identified at the threshold.
Extrapolation and smoothness

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

\[
\tau_{SRD} = E[Y_i(1) - Y_i(0)|X_i = c] = E[Y_i(1)|X_i = c] - E[Y_i(0)|X_i = c]
\]

- 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?

- 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.
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.

The discontinuity, except among municipalities with less than a 1996 electorate of about 10,000. The stability of the conditional expectation over such a large range of the data suggests that the treatment effect at $E_j = 40500$ may apply to municipalities far from the threshold.

Formal treatment effect estimates on null and blank votes—separately and together—are reported in the left panel of figure 7. Focusing on the local linear regression estimates, the effect of the shift in voting technology lowered null vote rates by an estimated 13.5 percentage points, blank votes by an estimated 10 percentage points, thus increasing the number of votes affecting political outcomes by about 23 percentage points. This number amounts to about a 34% increase in the size of the electorate casting valid votes. While null votes were somewhat more affected than blank votes, the similarity between the two estimates is surprising. A blank vote, in the Brazilian system, is supposed to be an affirmative choice intended by the voter. A null vote, on the other hand, is a residual category (an "undervote", to use the American parlance) for when the voter fails to register any preference at all. Thus, one might expect that electronic voting would affect null votes much more than blank votes, but these estimates belie that expectation. These estimates suggest that a large percent of blank votes were actually mistakenly cast or counted.

For comparison, treatment effect estimates on invalid votes for all other offices are reported in the right panel of figure 7. Electronic voting lowers invalid vote rates for all other offices, though estimates are smaller in magnitude.
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:
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.
- → 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

- 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 2016)

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:

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

- 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)
- 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: Monotonicity

There exists $\varepsilon$ 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 $c$,

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

Basically, in an $\varepsilon$-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.
- 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.
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:

<table>
<thead>
<tr>
<th>Discharge Type</th>
<th>Non HDC</th>
<th>HDC</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage Female</td>
<td>10.5</td>
<td>9.7</td>
<td>10.3</td>
</tr>
<tr>
<td>Mean Age at Release</td>
<td>28.9</td>
<td>30.7</td>
<td>29.3</td>
</tr>
<tr>
<td>Percentage Incarcerated for Violence</td>
<td>19.8</td>
<td>18.2</td>
<td>19.4</td>
</tr>
<tr>
<td>Mean Number Previous Offences</td>
<td>9.5</td>
<td>5.7</td>
<td>8.7</td>
</tr>
<tr>
<td>Recidivism within 12 Months</td>
<td>54.6</td>
<td>28.1</td>
<td>48.8</td>
</tr>
<tr>
<td>Sample Size</td>
<td>18,928</td>
<td>5,351</td>
<td>24,279</td>
</tr>
</tbody>
</table>

Panel B - Released +/- 7 Days of 3 Months (88 Days) Cu-off:

<table>
<thead>
<tr>
<th>Day of Release around Cut-off</th>
<th>- 7 Days</th>
<th>+ 7 Days</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage Female</td>
<td>11</td>
<td>10.2</td>
<td>10.3</td>
</tr>
<tr>
<td>Mean Age at Release</td>
<td>28.8</td>
<td>29.4</td>
<td>29.3</td>
</tr>
<tr>
<td>Percentage Incarcerated for Violence</td>
<td>17.1</td>
<td>19.7</td>
<td>19.4</td>
</tr>
<tr>
<td>Mean Number Previous Offences</td>
<td>9.1</td>
<td>8.6</td>
<td>8.7</td>
</tr>
<tr>
<td>Recidivism within 12 Months</td>
<td>56.8</td>
<td>47.9</td>
<td>48.8</td>
</tr>
<tr>
<td>Percentage Released on HDC</td>
<td>0</td>
<td>24.4</td>
<td>22</td>
</tr>
<tr>
<td>Sample Size</td>
<td>2,333</td>
<td>21,946</td>
<td>24,279</td>
</tr>
</tbody>
</table>
As discussed above, we need to illustrate the discontinuity of HDC treatment graphically and also continuity of covariates which could influence the recidivism outcome. Figure 1 begins this by plotting the proportion of prisoners discharged on HDC with respect to the length of their sentence **. After the 88 time limit the jump of 24.4 percent jump afore mentioned in proportion treated is clearly visible and highly significant.

Figure 1: Proportion Discharged on HDC by Sentence Length

Discontinuity in Proportion Discharged on HDC, Cut-off Point at 88 Days
Difference = .244 (.003)

**All the graphs are local polynomials with a 7 day bandwidths to be comparable to our chosen window around the threshold for RDD estimations.
Example: Early Release Program (HDC)

Figure 2: Mean Number of Previous Offence by Sentence Length

Figure 3: Mean Age at Discharge by Sentence Length
Figure 2 shows the mean number of previous offences by sentence length. Although the criminal history of prisoners is very different across the discharge period, the graph is very smooth around the HDC release threshold. This reinforces the validity of carrying out an RDD estimation of HDC as the number of previous offences could be strong selection criteria for scheme participation but is continuous around our assignment variable.

Figure 3 considers the mean age when released on sentence length. Again the graph is very continuous and the very small gap of the lines around the cut-off is not statistically significant. This again, with Figure 2, points to a relatively random distribution of observable characteristics around the threshold which further validates the use of RDD.

Finally, Figure 4 plots the rate of recidivism within 12 months of release by length of sentence. This graphical representation of the changes of our outcome variable of interests exhibits a striking jump around the 88 day threshold. This gap corresponds to a significant 8.9 percentage point lower re-offending rate of prisoners discharged one
Table 4: RDD Estimates of HDC Impact on Recidivism – Around Threshold

<table>
<thead>
<tr>
<th>Estimated Discontinuity of HDC Participation at Threshold (HDC⁺ – HDC⁻)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated Difference in Recidivism Around Threshold (Rec⁺ – Rec⁻)</td>
<td>-.089</td>
<td>-.059</td>
<td>-.044</td>
</tr>
<tr>
<td>Estimated Effect of HDC on Recidivism Participation (Rec⁺ – Rec⁻) / (HDC⁺ – HDC⁻)</td>
<td>-.366</td>
<td>-.268</td>
<td>-.181</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>PSM</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Sample Size</td>
<td>24,279</td>
<td>24,279</td>
<td>24,279</td>
</tr>
</tbody>
</table>

Dependent Variable = Recidivism Within 12 Months

Estimation on Individuals Discharged +/- 7 Days of 88 Days Threshold

Note: Robust standard errors in parenthesis. The estimation is based on individuals released between 89 and 180 days. The controls included in column (2) are: gender, age, number previous offences, month and year of release dummies, and the type of crime incarcerated for (8 types). The propensity score matching in column (3) is based on calculating propensity scores for each individual using the same variables as the controls in the previous model. Since controlling for observable characteristics does affect our estimate, we can go further and implement mix both PSM and RDD methodologies to obtain even more robust policy estimators. This is what we do in column (3) of Table 4. The main change now is that the difference in recidivism rates between prisoners discharged pre and post cut-off is significantly smaller at -4.4 percent. Consequently we can calculate that HDC participation reduces recidivism by about 18 percent ‡‡. This is very close to our preferred estimated impact from the OLS with PSM above and we therefore conclude that being released early on electronic monitoring from prison appears to reduce recidivism probability by between 18 and 19 percentage points.

‡‡ This is to our knowledge the first time that PSM and RDD have been combined to estimate a causal effect and there is therefore no simple methodology to obtain standard errors for this estimate. The reason is that we are not able here to run a local IV as before to generate the standard error. We however believe that the .181 coefficient is significant in view of the .044 standard errors in columns (1) and (2) and are working on a way to compute it precisely in the near future.
Example: Teamwork

- Qualified, Declined Invitation
- Superforecaster (accepted invitation)

Treatment
- Did Not Qualify for Super-Team Invitation
- Qualified/Received Super-Team Invitation

Year 2 Accuracy (Mean Std. Score) vs. Year 1 Accuracy (Centered Decision Score)
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
- 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

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
References


Approaches to Unmeasured Confounding

Natural Experiments

Traditional Instrumental Variables

Fun with Coarsening Bias

Modern Approaches to Instrumental Variables

Regression Discontinuity

Fun with Extremists

Fun With Weak Instruments

Appendix

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

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

- 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

- 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

Selection Bias.

- Districts choose extremist nominees because they prefer them.
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

Dem Primary $\cdots$ $M$ $\cdots$ Rep Primary
“Extremists” Defined

Dem Primary

M

Rep Primary

“Extremist”

“Moderate”
“Extremists” Defined

Dem Primary  \( M \)  Rep Primary

“Moderate”  “Extremist”
“Extremists” Defined

Dem Primary

Moderate? Extremist? Rep Primary

M

Stewart (Princeton)
“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

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%.
Quick Example: Robbie Wills vs. Joyce Elliott

Joyce Elliott: -0.33

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

\[ Y_{it} = \beta_0 + \beta_1 \text{Extremist Primary Win}_{it} + f(V_{it}) + \epsilon_{it} \]

\[ V_{it} \equiv \text{extremist candidate’s vote-share winning margin.} \]
0.4
0.6
0.8
−0.2 −0.1 0 0.1 0.2
N=233

General Election Vote Share

Extreme Candidate Primary Election Winning Margin

N=233
General Election Vote Share

Extreme Candidate Primary Election Winning Margin

N=233
Vote-share decrease: 

−0.09
Large Electoral Penalty to Nominating Extremist

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

Effect on Prob of Victory

5% Bandwidth
Local Linear
N=84

All Data
Cubic
N=253

IK Bandwidth=9.69%
Local Kernel
N=149

0

-1

-0.5

-1

Stewart (Princeton)

Week 11: Unmeasured Confounding

December 5 and 7, 2016
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?

![Graph showing win probability vs ideology](image-url)
How Does Penalty to Extremists Affect Roll-Calls?

![Graph showing Ideology vs Win Probability]

- Ideology
- Win Probability

Week 11: Unmeasured Confounding

December 5 and 7, 2016
Effect of Extremists on Roll-Call Voting

-1.0
-0.5
0.0
0.5
1.0

Terms Downstream

Democratic Primaries
Republican Primaries

More Conservative

More Liberal

Stewart (Princeton)
Effect of Extremists on Roll-Call Voting

-0.5
0.0
0.5
1.0
Terms Downstream
Democratic Primaries
Republican Primaries
0 1 2 3 4
More Conservative
More Liberal

Stewart (Princeton)
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

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

- Argument: Differential entry of extremist candidates forces voters to elect extremists.
Fun With Related Work


Approaches to Unmeasured Confounding

Natural Experiments

Traditional Instrumental Variables

Fun with Coarsening Bias

Modern Approaches to Instrumental Variables

Regression Discontinuity

Fun with Extremists

Fun With Weak Instruments

Appendix
(Thanks to Yuki Shiraito for sharing these slides with me)
Example: Endogeneity of Institution and Growth

Endogeneity Bias

- $Y_i$ and $D_i$ have direct causal effects on each other
- Ordinary least squares biased
Solution of Acemoglu et.al. (2001)

**Instrumental Variable Analysis**

- Exploiting exogeneity of $Z_i$
- Validity: Settlers chose to build stronger (non-extractive) institutions in places they wanted to live
- Exclusion: But early settler mortality rate does not affect GDP directly

Diagram:

- $Y_i$: GDP
- $D_i$: Institutions
- $Z_i$: Settler Mortality

Stewart (Princeton)  
Week 11: Unmeasured Confounding  
December 5 and 7, 2016
Addressing Endogeneity

Stewart (Princeton)

Week 11: Unmeasured Confounding

December 5 and 7, 2016
Addressing Endogeneity

\[ \begin{array}{|c|c|}
\hline
\text{OLS} & \text{TSLS} \\
\hline
\end{array} \]

Estimated Causal Effect

95% Confidence/Credible Interval

Bootstrapped 95% Confidence Interval

Stewart (Princeton)

Week 11: Unmeasured Confounding

December 5 and 7, 2016
Problem: Bootstrapped Confidence Interval

![Diagram showing estimated causal effects with OLS and TSLS methods, along with 95% confidence/credible intervals and bootstrapped 95% confidence intervals.](image-url)
Problem: Weak Instruments

- **Non-compliers**: For some countries, early settler mortality rate did not affect institutions.

- **Weak Instrument**: If many non-compliers in data,

\[
\hat{\beta}_{IV} = \frac{\text{Cov}(Y_i, D_i)}{\text{Cov}(D_i, Z_i)} \approx 0
\]

- Arbitrarily misleading estimates
Solution: Complier Instrumental Variable Estimation

- Want to focus on observations where $Z_i$ moves $D_i$, compliers
- Can’t observe who are compliers
- Estimate latent variable indicating compliance
- Strengthen the instrument through upweighting compliers
- Connect finite mixture modeling of compliance with weak instrument problem (Hirano et al. 2000)
IV as Simultaneous Equations

Standard model

\[
D_i = Z_i^\top \delta + X_i^\top \theta + \eta_i \quad \text{(First Stage)}
\]
\[
Y_i = D_i^\top \beta + X_i^\top \gamma + \epsilon_i \quad \text{(Second Stage)}
\]

where, for a simple random sample of \( i \in \{1, 2, \ldots, N\} \)

- \( Y_i \): Outcome
- \( D_i \): Endogenous
- \( Z_i \): Instrument
- \( X_i \): Covariates
- \( \epsilon_i, \eta_i \): Normal errors
- \( \text{Cov}(\epsilon_i, \eta_i) \neq 0 \)
CIV as Simultaneous Equations

CIV model

\[
\Pr(C_i = 1) = \Phi \left( W_i^\top \alpha \right) \quad \text{(Compliance Model)}
\]

\[
D_i = \begin{cases} 
\delta_0^C + Z_i^\top \delta + X_{i}^\top \theta + \eta_i; & C_i = 1 \\
\delta_0^{NC} + X_{i}^\top \theta + \eta_i; & C_i = 0 
\end{cases}
\quad \text{(First Stage)}
\]

\[
Y_i = D_i^\top \beta + X_{i}^\top \gamma + \epsilon_i 
\quad \text{(Second Stage)}
\]

where, for a simple random sample of \( i \in \{1, 2, \ldots, N\} \)

- \( C_i \): Indicator for complier (unobserved)
- \( W_i \): Covariates for compliance
- \( \Phi \): Normal CDF
- \( \epsilon_i, \eta_i \): Normal errors

Maximization not straightforward
- Gibbs sampler
- ECM algorithm in paper
Revisiting Acemoglu, Johnson, and Robinson (2001)

Causal effect of property rights on economic growth

- Outcome: 1995 GDP, logged per capita
- Endogenous variable: Risk of property expropriation
- Instrument: Mortality rate of European colonizers
- Covariates: Latitude (absolute value); former French colony (0/1) or British colony (0/1); proportion citizens who are Catholic, Muslim, and neither; whether the country has a French legal origin (0/1)
- $N=64$

CIV for

- Strengthening a weak instrument
- Characterizing compliers
Causal Estimates

- OLS
- TSLS

Estimated Causal Effect
95% Confidence/Credible Interval
Bootstrapped 95% Confidence Interval

Stewart (Princeton)
Week 11: Unmeasured Confounding
December 5 and 7, 2016
Causal Estimates

Estimated Causal Effect

95% Confidence/Credible Interval

Bootstrapped 95% Confidence Interval

OLS

TSLS

CIV
Causal Estimates

Estimated Causal Effect

95% Confidence/Credible Interval

Bootstrapped 95% Confidence Interval

OLS
TSLS
CIV
RANK
JACK
Compliance, by Colonizer

Stewart (Princeton)
Week 11: Unmeasured Confounding
December 5 and 7, 2016
They propose **CIV** as a solution to the weak instrument problem

- **Weak instruments**
  1. Many non-compliers in data
  2. Arbitrarily misleading estimates

- **Complier Instrumental Variable estimation**
  1. Estimates latent compliance status
     → Characterize who are the compliers
  2. Upweights the compliers
     → Strengthens the instrument
Approaches to Unmeasured Confounding

Natural Experiments

Traditional Instrumental Variables

Fun with Coarsening Bias

Modern Approaches to Instrumental Variables

Regression Discontinuity

Fun with Extremists

Fun With Weak Instruments

Appendix
Proof of the LATE theorem

- Under the exclusion restriction and randomization,
  \[ E[Y_i|Z_i = 1] = E[Y_i(0) + (Y_i(1) - Y_i(0))D_i|Z_i = 1] = E[Y_i(0) + (Y_i(1) - Y_i(0))D_i(1)] \] (randomization)

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

- The third equality comes from monotonicity: with this assumption, \( D_i(1) < D_i(0) \) never occurs.
\[ E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0] = E[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)] \Pr[D_i(1) > D_i(0)] \]

- We can use the same argument for the denominator:

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

- Dividing these two expressions through gives the LATE.