DS-UA 201: Causal Inference: Regression Discontinuity

Parijat Dube

New York University Center for Data Science

August 15, 2022

Acknowledgement: Slides including material from DS-US 201 Fall 2021 offered by Marco Morucci.

So far: Identification under unobserved confounding

In the last few weeks we have been focusing on identifying treatment effects under some (structured) forms of **unobserved confounding**

- DiD: Unobserved confounding is time-invariant
- IV: Unobserved confounding affects the treatment but not the instrument

We have seen how **linear regression** is a standard tool in these settings.

- Sometimes just a short hand for difference-in-means estimators (works under mild assumptions)
- Sometimes an actual outcome model (stronger assumptions needed!)

Today: Regression Discontinuity Designs

Regression Discontinuity Designs (RDD) is another approach to identification when there is unobserved confounding.

- **Goal**: Use knowledge about particular treatment assignment mechanisms to find a subset where treatment is "quasi-randomly" assigned.
- **Intuition**: Exploit the fact that treatment is sometimes assigned based on a **cut-off** or threshold value of a continuous running variable
  - Merit aid scholarships awarded based on a test-score cut-off
  - Geographic boundaries affect policy implementation.
  - Candidates receiving a plurality of the vote get elected.

Same as last weeks, we allow unobserved confounding, but we restrict its structure with assumptions.

Regression Discontinuity Designs (RDD)

- **Regression**: We're working with a conditional expectation (regression) function (CEF) of the outcome given a running variable.

- **Discontinuity**: There is a *jump* in the CEF that is driven by treatment-assignment at some cut-off
- **Design**: The assumptions are informed by some substantive knowledge of a particular situation

Quasi-randomization at a cut-off

Key intuition: The continuous variable is a confounder, but units that are close to the cut-off are very similar in characteristics **except** for the fact that some get treatment vs. control.

- Not too much of a difference between candidates receiving 49.9% of the vote and candidates receiving 50.1% of the vote. Except one wins and the other loses.

Key assumption: The relationship between the continuous variable and outcome would be **continuous** but for the discontinuity in treatment assignment.

- Units are not able to control their score precisely such that near the discontinuity, the assignment is "as-good-as-random"

Example: Incumbency advantage (Lee, 2008)

How much does being an incumbent boost a candidate's probability of winning an election?

Problem: Hard to disentangle effect of incumbency over other district characteristics.

- Is a Democrat (or Republican) more likely to hold on to a seat because they're incumbents or because they align with district preferences?

Lee (2008) exploits the fact that U.S. congressional elections are "first-past-the-post" – the plurality vote-getter wins the seat.

- Districts where the Democratic candidate **won** by a tiny margin and ones where they **lost** by a tiny margin are very similar...except for the fact that the incumbent party is different!

Lee 2008: Research Design

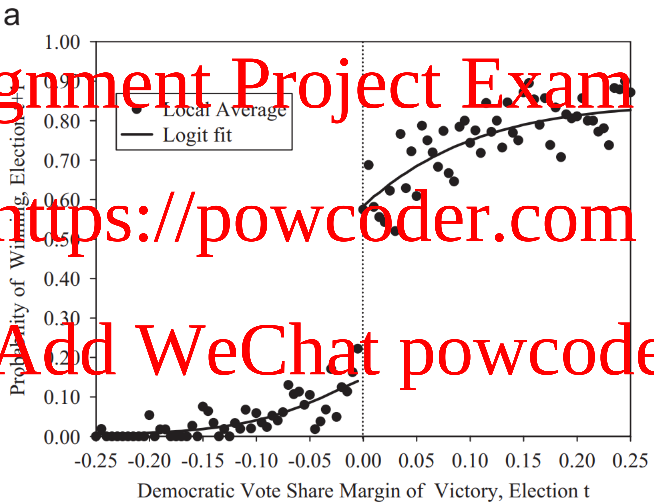
Design: Examine the probability of Democratic victory in an election at time $t + 1$ conditional on the Democratic margin-of-victory in an election at t.

- Only look at those districts where democrats won or lost by a small margin in the previous election.

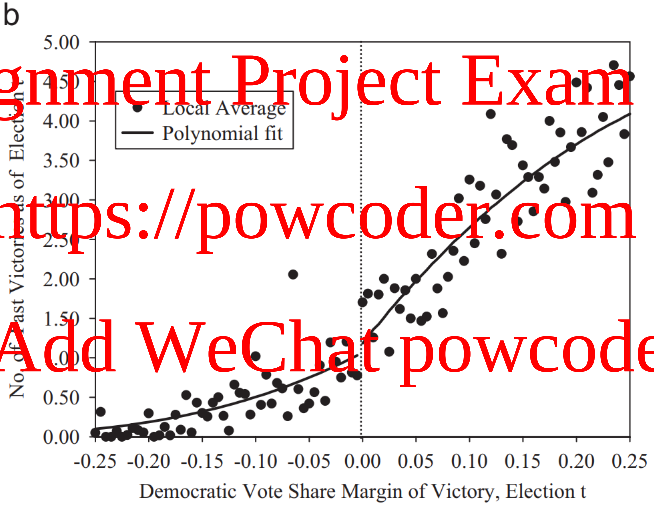
Placebo test: If there was something particular about those districts where democrats win at $t + 1$ then it should have been affecting victory in previous times as well.

- If there is no discontinuity in previous margin of victory, then there is evidence that there is no confounding.

Example: Incumbency advantage (Lee, 2008)



Example: Incumbency advantage (Lee, 2008)



RDD Setup

We will be working with the following setting:

- Binary treatment $D_i \in \{0, 1\}$
- Potential outcomes $Y_i(1), Y_i(0)$.
- Observed outcomes/consistency:
  $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$

We also have $X_i \in \mathbb{R}$, which is a continuous variable that affects treatment assignment.

- Sometimes referred to as the running or forcing variable.

Assumption 1: Treatment assignment

Sharp RD – Treatment assignment is perfectly determined by the value of the forcing variable X_i and the threshold c

Assignment at the discontinuity

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

Units with X_i above the cut-off/threshold c get the treatment.
Units with X_i below the cut-off get control.

Assumption 2: Continuity in potential outcomes

Assumption 2: The conditional expectation of the potential outcomes given X_i are continuous.

Continuity

$$E[Y_i(0)|X_i = x] \text{ and } E[Y_i(1)|X_i = x]$$

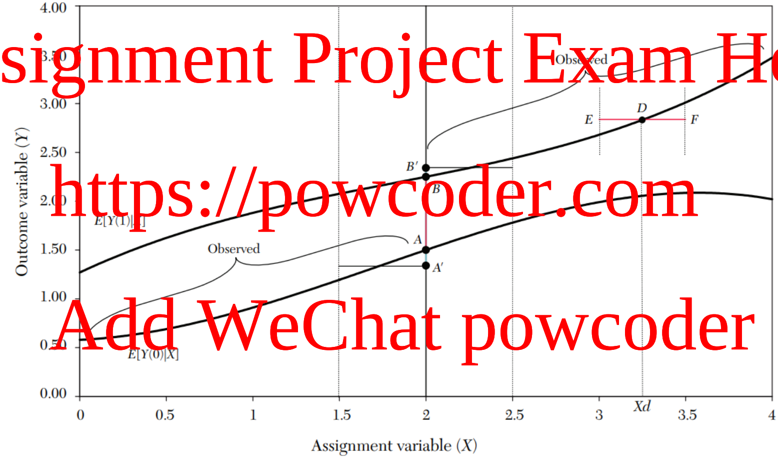
are continuous in x

In other words:

$$\lim_{x \to c} E[Y_i(d)|X_i = x] = E[Y_i(d)|X_i = c]$$

As x gets infinitesimally closer to c, we will have: $E[Y_i(d)|X_i = x] = E[Y_i(d)|X_i = c]$

Visualizing the regression functions



Identification using limits

Our identification strategy leverages the fact that the limit of the CEF as we approach the discontinuity will be different depending on whether it is from the **right** vs. **left**.

$$\tau = \lim_{x \to c+} E[Y_i|X_i = x] - \lim_{x \to c-} E[Y_i|X_i = x]$$

- The limit from the **right** ($x \to c+$) identifies $E[Y_i(1)|X_i = c]$ (all values of $X_i > c$ are "treated")
- The limit from the **left** ($x \to c-$) identifies $E[Y_i(0)|X_i = c]$ (all values of $X_i < c$ are "control")

Identification using limits

How do the assumptions get us identification?

$$
\begin{aligned}
E[Y_i(0)|X_i = c] &= \lim_{x \to c-} E[Y_i(0)|X_i = x] \\
&= \lim_{x \to c-} E[Y_i(0)|D_i = 0, X_i = x] \\
&= \lim_{x \to c-} E[Y_i|X_i = x]
\end{aligned}
$$

Same intuition for the limit from the right ($x \to c+$) for identifying $E[Y_i(1)|X_i = c]$

RDD and conditional ignorability

RDD implicitly makes a **conditional ignorability** assumption. We know:

$$\Pr(D_i = 1) = \mathbb{1}(X_i > c),$$

therefore:

$$\begin{aligned}\Pr(D_i = 1 | Y_i(d), X_i = x) &= \begin{cases} 1 & x > c \\ 0 & x < c \end{cases} \\ &= \Pr(D_i = 1 | X_i = x)\end{aligned}$$

Treatment assignment is **independent** of the potential outcomes because it is **perfectly determined** by X!

Extrapolation

However, **positivity** is violated!

- Probability of treatment is 0 for units with $X_i < c$ and 1 for units with $X_i > c$.
- There is no covariate overlap between treated and untreated units, i.e., all treated units have $X_i > c$ and all control units have $X_i < c$.
- There are no units such that $X_i = c$

RDD is about **extrapolating** from the treated observations and control observations to a *common* value of X_i (the discontinuity).

- "what if" there were units exactly at the threshold?

Identifying the "local" treatment effect

RDD identifies a **local** treatment effect – the treatment effect for units **at** the discontinuity.

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

How to interpret this?

- Can think of the RDD as a "locally" randomized experiment (units close to, but above the cut-off are comparable to those close to, but below the cut-off)
- Implicitly making a no effect heterogeneity assumption to generalize (at least for units close to the discontinuity)
- How far can we extrapolate? Depends on how much treatment effects vary.

Estimation challenges

Key Idea: With infinite data, we can get arbitrarily close to the true ATE at the discontinuity. More and more observations with X_i really close to c

- However, in actual datasets, we have to *extrapolate* to the discontinuity using observations that might be kind of far away.

Bias-Variance Trade-off.

- Using observations that are far from the discontinuity increases bias (since $Y_i(1)$ and $Y_i(0)$ are associated with X_i), but reduces variance (more observations).
- Using only "close" observations reduces bias but increases variance (fewer observations).

Binned scatterplots

The best RDDs are the ones you can **see**.

- Make a plot of the average of Y_i within "bins" of X_i.
- Is there an obvious gap near the means around the cut-point?
- Is the conditional expectation changing comparatively smoothly for X_i far from the cut-point?
- Do this for covariates/placebo outcomes as well – we should *not* see a discontinuity around the cut-point.

Illustration: Lee (2008) Election RDD

We'll illustrate this using the Lee (2008) election dataset

- X: Democratic margin of victory in time t
- Y: Democratic vote share in time $t+1$
- D: Victory in time t (margin > 0).

Illustration: Lee (2008) Election RDD

```r
## We're just using this for the Lee (2008) dataset
library(rddtools)
library(estimatr)
library(tidyverse)

## Load the Lee (2008) data
data(house)

# What does it look like?
# x (vote share at time t-1)
# y (vote share at time t)
> head(house)
        x      y
1  0.1049 0.5810
2  0.1393 0.4611
3 -0.0736 0.5434
4  0.0868 0.5846
5  0.3994 0.5803
6  0.1681 0.6244
```

Illustration: Lee (2008) Election RDD


```r
# The regular scatterplot is really noisy -- let's bin
    the points instead
bin_scatter <- ggplot(aes(x=x, y=y), data=house) +
  stat_summary_bin(fun.y='mean', bins=200,
                   size=2, geom='point') +
  geom_vline(xintercept=0, col="red", lty=2) +
  xlab("Democratic vote share margin of victory,
     Election t") +
  ylab("Democratic vote share, Election t+1") +
  theme_bw()
```

Illustration: Lee (2008) Election RDD

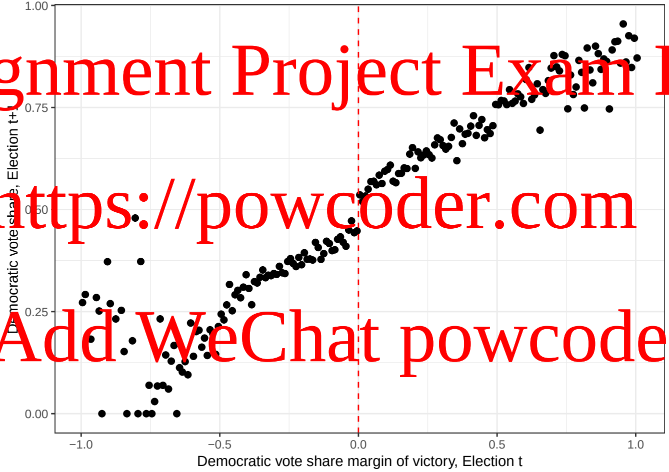
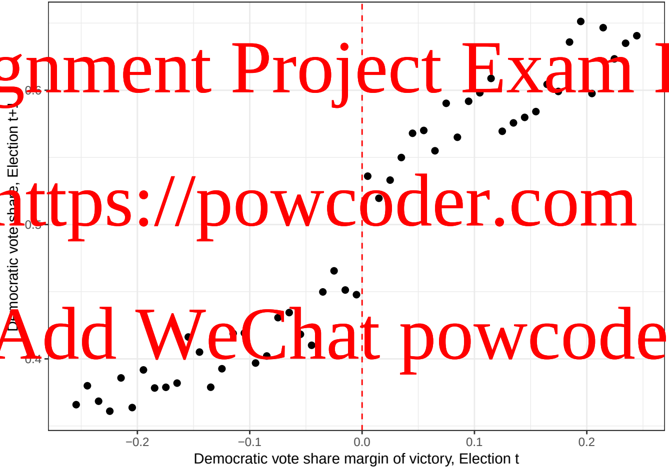


Illustration: Lee (2008) Election RDD

Now change to outcome variable:

- X: Democratic margin of victory in time t
- Y: Democratic vote share in time $t + 1$
- D: Victory in time t (margin > 0).

Illustration: Lee (2008) Election RDD



Local linear regression

How do we get **point estimates** for the TE?

- **Goal**: Estimate $\lim_{x \to c+} E[Y_i|X_i]$
- Fit a **model** to the treated units and get the prediction at the cut-point.
  - This requires assumptions on the form of Y!
- To reduce dependence on getting the correct model, use only units with X_i close to c (within some "bandwidth" h)
- Use a regression to account for how $E[Y_i|X_i = x]$ changes with x (even for those "close" units).

Local linear regression

- Let $\hat{\mu}_+(x)$ denote the predicted value from a regression of Y_i on X_i fit to observations within the bandwidth above the cut-point $(c, c+h]$.

- Let $\hat{\mu}_-(x)$ denote the predicted value from a regression of Y_i on X_i fit to observations within the bandwidth below the cut-point $[c-h, c)$.
- Our estimate of the ATE is the difference between the predictions at c.

$$\hat{\tau}_{\text{RD}} = \hat{\mu}_+(c) - \hat{\mu}_-(c)$$

Local linear regression

Can get this (+ valid SEs) from a single regression with interactions:

- Subset the data to only observations with X_i within h of the cutpoint c (the "close" observations). Then fit:

$$Y_i = \alpha + \tau D_i + \beta(X_i - c) + \gamma(X_i - c)D_i + \epsilon_i$$

- $\hat{\tau}$ is our estimate of the RD effect – usual methods for getting SEs.

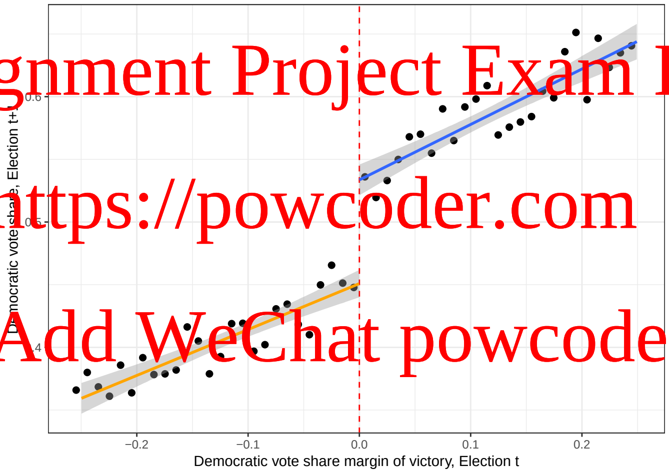
Illustration: Lee (2008) Election RDD

```
# generate a treatment indicator
house$d <- as.integer(house$x > 0)

# Subset to the close observations
house_close <- subset(house, x>-.25&x<.25)

# Fit the regression model w/ interaction
rd_reg <- lm_robust(y ~ d + x + d*x, data=house_close)
> rd_reg
             Estimate Std. Error t value Pr(>|t|) CI
                 Lower CI Upper   DF
(Intercept)    0.4509    0.00558   80.82 0.00e+00
    0.4399    0.4618 2757
d              0.0827    0.00838    9.87 1.31e-22
    0.0663    0.0991 2757
x              0.3665    0.04135    8.86 1.37e-18
    0.2854    0.4476 2757
d:x            0.0760    0.06288    1.21 2.27e-01
    -0.0473    0.1993 2757
```

Illustration: Lee (2008) Election RDD

```r
# Add it to the plot
# Scatterplot w/ regression
bin_scatter_close_reg <- ggplot(aes(x=x, y=y), data=
    house_close) +
  stat_summary_bin(fun.y='mean', bins=50,
                   size=2, geom='point') +
  geom_vline(xintercept=0, col="red", lty=2) +
  geom_smooth(data=subset(house_close,d==1), formula= y
      ~ x, method="lm_robust") +
  geom_smooth(data=subset(house_close,d==0), formula= y
      ~ x, method="lm_robust", col="orange") +
  xlab("Democratic vote share margin of victory
    Election t") +
  ylab("Democratic vote share, Election t+1") +
  theme_bw()
```

Illustration: Lee (2008) Election RDD



Summary

Today we have introduced **Regression Discontinuity Designs**

- Assumes that treatment is assigned past a threshold c on a continuous variable X
- Identification at the threshold possible with limits
- Inference with local linear regression

Up next

- Implementing and making RDD plots
- Fuzzy RDD (when the forcing variable increases probability of treatment in a discontinuous way).
- Diagnosing RDD assumptions