Review Session 8
DiD, IV, RD

Ben Berger

March 24, 2023

Today

- Review Difference-in-Differences Model
    - Tenancy Reform Conceptual Exercise
- Instrumental Variables
    - Example on colonoscopy invitations
    - R exercise on cigarette consumption
- Regression Discontinuity
    - Conceptual Exercise on British parliamentary candidates
    - R tutorial on RD estimation and plotting

Difference-in-Differences

What's the Diff-in-diff?

- *Very* popular research method using comparability of trends in two groups.
- Basic 2×2 DiD:

$$Y_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 Post_t + \beta_3 Treat_i \times Post_t + u_{it}$$

Set Up/What You Need

- Observations of units before and after a treatment in which only some were exposed.
- e.g. two municipalities, two drug classes, two school districts, etc.

Mechanics

- For 2×2 DiD, estimate using linear regression with dummy-dummy interactions!
- For models with more time periods, can use many time and group fixed-effects.

Difference-in-Differences

	Before	After	Difference
Control	β_0	$\beta_0 + \beta_2$	β_2
Treatment	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_2 + \beta_3$
Difference	β_1	$\beta_1 + \beta_3$	β_3

Parallel Trends Assumption

Assumption Needed for Causal Interpretation

- Let $Y_0(t)$ be the untreated potential outcome in period t:

    Parallel Trends $\iff E[Y_0(1) - Y_0(0)|Treat = 0] = E[Y_0(1) - Y_0(0)|Treat = 1]$

- The change in outcome for the control group is equal to the change that the treatment group would have experienced in the absence of treatment.
- Under this assumption, we attribute any observed differential increase as the causal effect of the policy.
- Treatment and control may vary in **level** of Y, but any difference in **trends** must be attributable to receiving treatment.

Parallel Trends Assumption

Can we test for parallel trends?

- We can't test whether treated units would have counterfactually trended parallel to non-treated units.
- The next best thing is to test whether the groups have parallel trends before treatment (AKA pre-trends).
- If the groups didn't follow parallel trends before treatment, they probably wouldn't have after either.

Difference-in-Differences

Variations on Difference-in-differences:

- Static: Estimate one time-invariant treatment effect

$$Y_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 Post_t + \beta_3 Treat_i \times Post_t + u_{it}$$
$$Y_{it} = \alpha_i + \lambda_t + \beta Treat_i \times Post_t + u_{it}$$

- Dynamic: Estimate effect for each time period

$$Y_{it} = \beta_0 + \sum_{k \neq 1} \beta_k Treat_i \times 1\{k = t\} + \alpha_i + \gamma_t + u_{it}$$

- Staggered Diff-in-diff:
    - Many papers estimate DiD with differences in treatment timing.
        - e.g. states adopt minimum wages at different times
    - Formerly common to estimate these with two-way fixed effects regression (individual + time dimensions).
    - But new literature points out this is biased (Goodman-Bacon 2021) and provides alternative estimators (Sun and Abraham 2020; Callaway and Sant'Anna 2021).

DiD Exercise: Tenancy Reform

Tenancy reform in West Bengal
   In the period before Operation Barga, agricultural productivity was growing at almost identical rates in the two states... Between 1969 and 1978, a period covering the decade before Operation Barga [a program that gave sharecroppers protection from eviction], rice yields increased by 9.3 percent in West Bengal and by 11 percent in Bangladesh. In the period after Operation Barga was introduced (1979-93), rice yields in West Bengal increased by 69 percent compared to 44 percent in Bangladesh... In the post Operation Barga period, rice yields in West Bengal are substantially higher in all years except for 1981 and 1982, when West Bengal experienced two successive years of severe droughts. (Benarjee, A., Gartler, P. and Ghatak, Maintreesh. (2002))
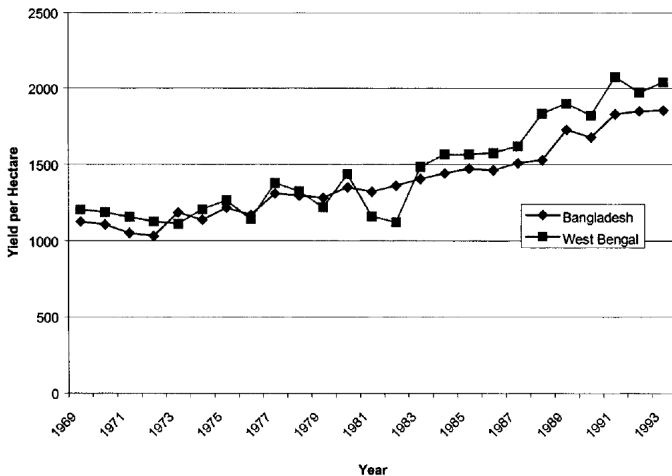
DiD Exercise: Tenancy Reform

Fig. 4.—Rice yield in West Bengal and Bangladesh, 1969–93

DiD Exercise 1: Tenancy Reform

Questions

1. Does the pre-intervention data help the authors to make a more favorable case for the plausibility of the parallel trend assumption? Why yes or why not?
2. Do the 1981-82 droughts makes the case for parallel trends weaker?
3. How would knowing the following (made up) facts affect your confidence in the difference-in-differences research design?
   1. Different varieties of rice are cultivated in West Bengal and Bangladesh.
   2. A major pro-market reform was implemented in West Bengal at the same time the tenure reform was implemented.
   3. Rainfall is more volatile from year to year in West Bengal than in Bangladesh.
   4. Consumption of rice has steadily fallen in West Bengal from the late 60's.

IV Intuition

- Recall that the issue with OLS estimation is that variation in the variable of interest may be related to other causes of Y.
- The core idea of IV is to find something that causes variation in your variable of interest that is **unrelated to all other causes of Y**.
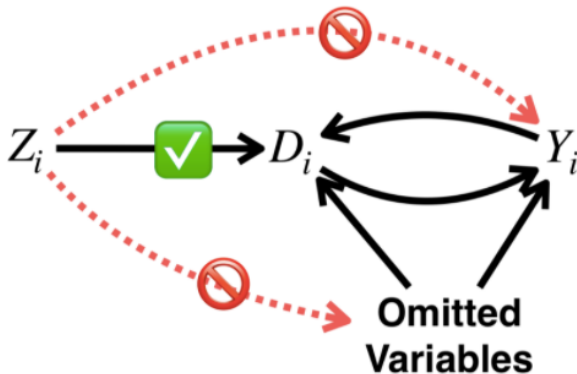
Bretthauer et al. 2022

Clinical Problem

Although colonoscopy is widely used as a screening test to detect colorectal cancer, it is more invasive and requires more resources than fecal occult blood tests and sigmoidoscopy. To determine whether its benefits outweigh these costs, additional high-quality data are needed from randomized trials.

Clinical Trial

Design: A multinational, pragmatic, randomized trial assessed the 10-year effects of population-based colonoscopy screening on the risks of colorectal cancer and related death.

Intervention: 84,585 men and women 55 to 64 years of age who lived in Poland, Norway, and Sweden and who had not previously undergone screening were randomly assigned either to receive an invitation to undergo onetime colonoscopy screening (invited group) or to receive no invitation or screening (usual-care group). The primary end points were the risks of colorectal cancer and related death after median follow-ups of 10 years and 15 years.

Bretthauer et al.

- Z: Colonoscopy invitation
- D: Colonoscopy receipt
- Y: Colorectal cancer, CRC death

How can we interpret the IV assumptions in this context?

- **Relevance:** Colonoscopy invitation affects colonoscopy receipt →Information about a procedure makes you more likely to get it
- **Exogeneity:** no paths between instrument and outcome other than through the endogenous variable. Two sub-conditions:
    - Direct: instrument cannot effect outcome directly →invitation doesn't have therapeutic properties, it's just a piece of paper.
    - Indirect: the instrument is independent of other causes of the outcome →receiving the invitation doesn't change health behaviors.
- **Monotonicity:** Let's circle back to this…

Populations in an IV Study

LATE framework developed by Angrist and Imbens (1994) to interpret binary IV estimates. With a binary instrument, there are 4 types of individuals:

- **Always-takers** are people who *would* take the treatment no matter what.
  →Get a colonoscopy regardless of invitation.
- **Never-takers** are people who *would not* take the treatment no matter what.
  →Never get a colonoscopy.
- **Compliers** are people who *would* take the treatment if encouraged to do so by the instrument but *would not* take the treatment if not encouraged to do so. →Get a colonoscopy only if invited.
- **Defiers** are people who *would not* take the treatment if encouraged to do so by the instrument and *would* take the treatment if not encouraged to do so.
  →Get a colonoscopy if not invited, don't get a colonoscopy if invited.

Monotonicity Assumption

Monotonicity:

- Effect of instrument on endogenous variable is in one direction →Invitation made no one less likely to get a colonoscopy, i.e. there are no defiers.

Interpretation:

- If monotonicity holds, $\hat{\beta}^{IV}$ is the Local Average Treatment Effect →the effect of colonoscopy on health measures for individuals who get a colonoscopy if invited and don't otherwise (the compliers).

IV Estimation: Two Stage Least Squares

Two equivalent ways to conduct IV regression:

With predicted values:

1. Estimate the first stage regression: $D = \alpha_0 + \alpha_1 Z + v$
2. Calculate the predicted values, $\hat{D}$
3. Using these $\hat{D}$, estimate the second stage regression: $Y = \beta_0 + \beta_1 \hat{D} + u$

As the ratio of two coefficients:

1. Run the first stage regression: $D = \alpha_0 + \alpha_1 Z + v$
2. Run the "reduced form" regression: $Y = \gamma_0 + \gamma_1 Z + \eta$
3. $\hat{\gamma}_1 / \hat{\alpha}_1$ is the unbiased estimate of the effect

IV Pros and Cons

What is so great about IVs?

- If you can identify some exogenous sources of variation that drive the treatment you can identify treatment effects even with unmeasured confounding on the treatment-outcome relationship!

What is less amazing?

- That's a big "if".
    - You need to be able to argue that instruments only impact the outcome through the explanatory variable. Good instruments are in high demand and low supply.
    - IV can be useful in RCTs if there is attrition. Allows estimation of causal effect of treatment not just randomization.
- IV identifies the LATE. While this might be a policy-relevant treatment effect and of interest in itself, but you will need to argue about its applicability to other groups.

Exercise 1: Cigarette Consumption (Stock and Watson)

Suppose we were interested in the impact of price on cigarette demand. Increasing cigarette prices might disincentivize smoking, but could also just impose a higher burden on smokers without stopping smoking.

Load packages and cigarette data using the following:

```
library(tidyverse)
library(haven)
library(fixest)

cigarettes <- read_dta("cigarettes.dta")
```

Exercise 1: Cigarette Consumption (Stock and Watson)

Panel data on 48 US states from 1985–1995.

Variables:

- packs: packs of cigarettes sold per capita
- price: average price per pack including sales tax
- salestax: average sales tax net of other excise taxes
- state: hopefully self-explanatory
- year: definitely not month or day

Estimate the following model

$$\log(\text{Packs}_{st}) = \beta_0 + \beta_1 \text{Price}_{st} + u_{st}$$

1. Using OLS.
2. Using state and year fixed-effects.
3. Instrumenting for sales tax.

In all specifications, cluster standard errors at the state level.

To estimate the regression of Y on D, instrumenting for Z (and including no controls) run the following: `feols(Y ~ 1 | D ~ Z, data)`.

Exercise 1: Cigarette Consumption (Stock and Watson)

```r
# OLS
fit_e1_ols <- feols(log(packs) ~ price,
                    cigarettes,
                    cluster = ~state)

# Two-way fixed effects
fit_e1_twfe <- feols(log(packs) ~ price | state + year,
                     cigarettes,
                     cluster = ~state)

# IV
fit_e1_iv <- feols(log(packs) ~ 1 | price ~ salestax,
                   cigarettes,
                   cluster = ~state)
```

Exercise 1: Cigarette Consumption (Stock and Watson)

Dependent Variable:		Log(Packs)	
	OLS	TWFE	IV
Model:	(1)	(2)	(3)
Variables			
Price	-0.0037***	-0.0054***	-0.0041***
	(0.0002)	(0.0008)	(0.0006)
Fixed-effects			
State		Yes	
Year		Yes	
Fit statistics			
Observations	96	96	96
F-test (1st stage), Price			117.30

Clustered (State) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Exercise 1: Cigarette Consumption (Stock and Watson)

Using IV, we've estimate the SRF:

$$\log(\widehat{\text{Packs}_{st}}) = 5.26 - 0.0041 \times \text{Price}_{st}$$

Interpret this IV estimate. You can use causal language.

A one dollar increase in the price of a pack of cigarettes leads to a 0.4 percent decrease in packs sold per person.

What assumptions must hold for this to be a causal effect?

Exercise 1: Cigarette Consumption (Stock and Watson)

Assumption 1 – Relevance: Does the instrument (sales tax) impact the endogenous variable (price)?

We can test this assumption by regressing price on sales tax.

```
feols(price ~ salestax, cigarettes, cluster = ~state)

OLS estimation, Dep. Var.: price
Observations: 96
Standard-errors: Clustered (state)
            Estimate Std. Error t value  Pr(>|t|)
(Intercept) 101.51140    5.209109 19.4873 < 2.2e-16 ***
salestax      7.43276    0.604781 12.2900 2.7548e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 29.1   Adj. R2: 0.550397
```

Exercise 1: Cigarette Consumption (Stock and Watson)

Assumption 2 – Exogeneity: Does the instrument (sales tax) impact the outcome ($\log(\text{packs})$) only through the endogenous variable (price)?

This might be violated if, for example, the sales tax causes prices of other goods to increase, decreasing consumers' ability to pay for cigarettes.

Unfortunately, we can't test this assumption directly unless we have multiple instruments.

Exercise 1: Cigarette Consumption (Stock and Watson)

Assumption 3 – Monotonicity: Does the instrument (sales tax) impact the endogenous variable (price) in only one direction?

Probably. It seems unlikely that an increase in sales tax would ever lead to a decrease in price including tax.

Unfortunately, we can't test for monotonicity either because we would need to know states' counterfactual prices given different sales tax regimes.

Regression Discontinuity Designs

RD is a popular research design because:

- It is intuitive and easy to visualize
- It occurs whenever there is a rule determining treatment assignment, which is very common in policy contexts (e.g. education, administrative programs, elections, etc.)

There are tradeoffs, however:

- Pro: high internal validity $\rightarrow$ has been shown to reproduce an experimental results
- Con: low external validity $\rightarrow$ effect only identified for observations *at* the cutoff, everything else is extrapolation.

Regression Discontinuity Designs

What can we do with our data to see if a RD makes sense?

- Plot *treatment* as a function of the running/forcing variable.

- When $\Pr(\text{Treatment})$ goes from $0 \rightarrow 1$ at the cutoff we have a Sharp RD. [1]

Forcing variable and Treatment Status

---
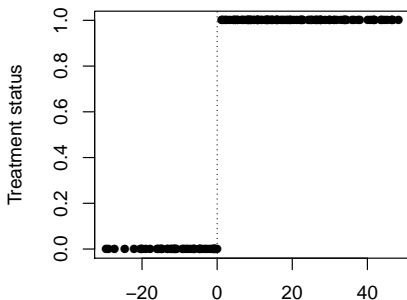
[1]You can also have Fuzzy RD where $\Pr(\text{Treatment})$ between 0 and 1, but can't estimate with OLS. Instead use IV, which scales the discontinuity by a first stage.

RD Assumptions & Interpretation

What are the RD assumptions?

- **Continuity** of average potential outcomes.
  - $E[Y_1|X]$ and $E[Y_0|X]$ are continuous at $X = c$.
- We can illustrate these potential outcomes graphically. Suppose X reflects a pre-test score and Y reflects a post-test score and the treatment is an intervention to help struggling students.
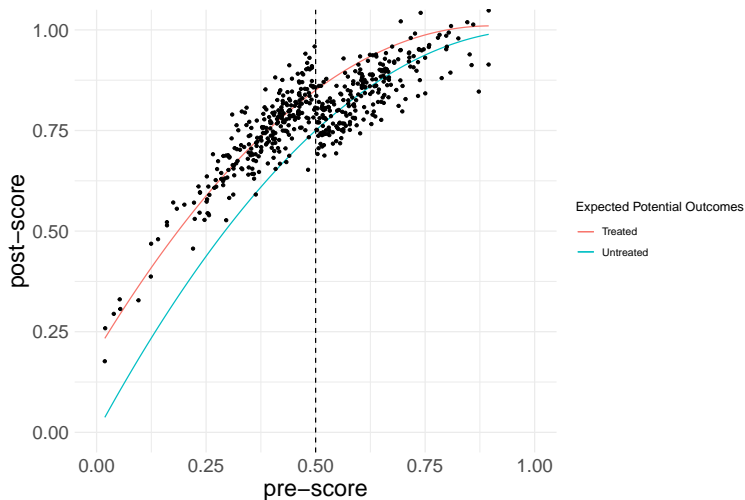
Interpretation

- RD estimates are only identified for observations at the cutoff; the further you get away from the cutoff, the more you need to extrapolate.
- Therefore, we get the Local Average Treatment Effect ("local" meaning for those at the cutoff)
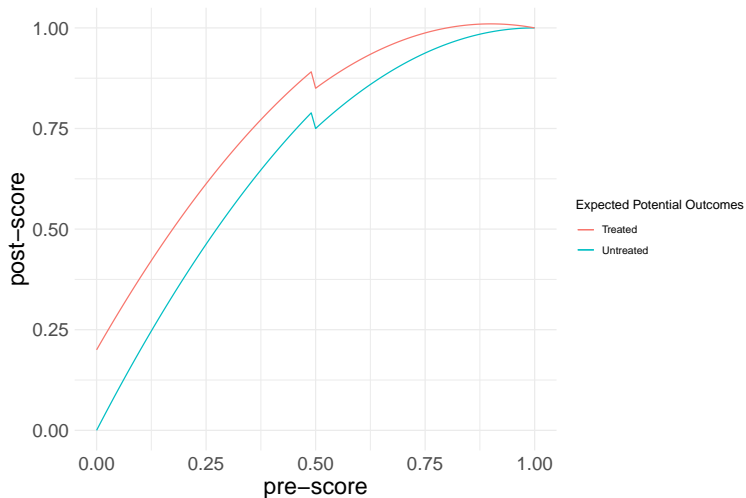
RD Continuity
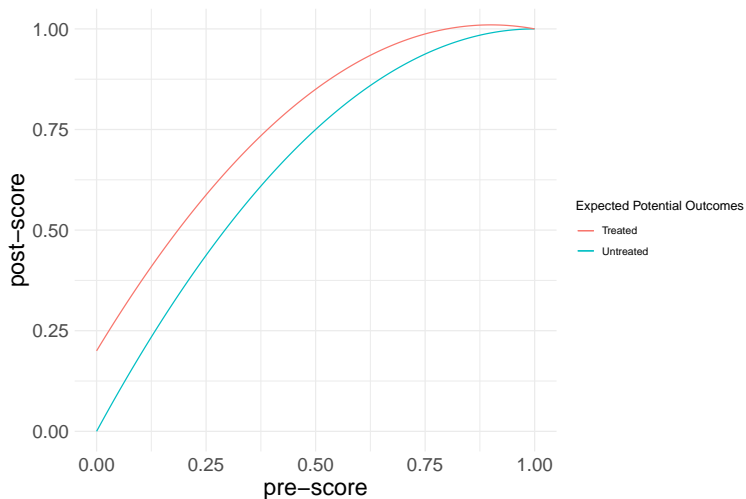
RD Continuity

RD Continuity Violation

Suppose that below the threshold (0.5), students are placed on academic probation and banned from participating in sports. Moreover, suppose that sports on average negatively affect performance on the post-test, perhaps because it takes away from possible time to study.

What do the potential outcomes of the *teaching intervention alone* look like now?

RD Continuity Violation

RD Continuity Violation

- Whether or not students receive the teaching intervention, if they fall below 0.5, they are banned from sports, increasing grades on average.
- Thus the potential outcomes for the teaching intervention are discontinuous at the cutoff.
- In this case RD does reflect the combined impact of the intervention AND academic probation on post-test score rather than the teaching intervention alone.
    - This total effect could be of interest, but the effects alone would probably be of greater interest to an education policy researcher.
- **Question:** What can we say about the differential effects of the teaching intervention on high performers (pre-score ≥ 0.75) and very low performers (pre-score ≤ 0.25).
    - Answer: Unfortunately nothing! RD estimates a **local** average treatment effect (LATE).
    - This makes sense though because we don't have any variation in treatment status for these individuals.

Analysis & Outcomes

Centering: By recentering our running variable so that it is equal to zero at the threshold, the coefficient on the treatment indicator becomes the RD estimand.

Formally: where D_i is treatment, X_i is the recentered running variable, and Y_i is outcome

- No change in slope: $Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + u$
- Change in slope: $Y = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 D_i \times X_i + u$
- You can also add polynomial terms to allow for curvature in running variable.
- Ultimately we care about β_1, the treatment effect at the cutoff.

Checking for Robustness

RDs are awesome because they have high internal validity and because they are easy to visualize. This makes it (somewhat) easier to assess the plausibility of their identification assumption: continuity.[2]

Some common tests and concerns:

- **Sorting:** If units can sort at the cutoff, then units just over and below the cutoff are no longer comparable.
- **Non-linearity:** Are results sensitive to including higher-order polynomials (quadratics, cubics, etc)?
- **Bandwidths:** Are results sensitive to changes in the window around the cutoff?
- **Placebos:** Do jumps occur where we would not expect them?
- **Balance:** We want continuity in covariates and the running variable around the threshold c.

---

[2]Although we can never "test" an assumption about potential outcomes because we can only ever observe one.
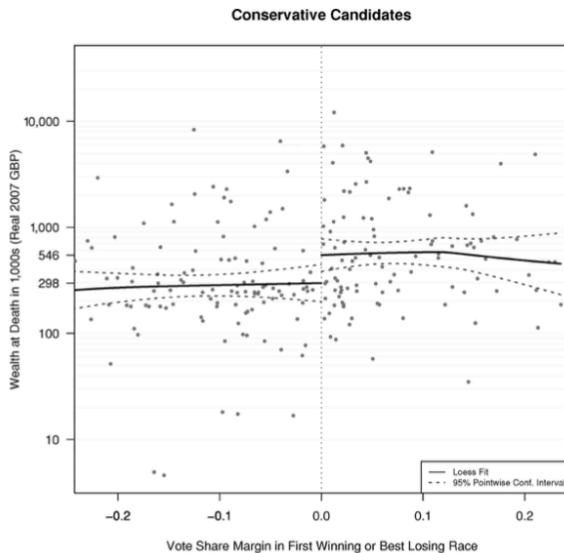
Exercise 2: Assessing Discontinuities

MPs for Sale? Returns to Office in Postwar British Politics

ANDREW C. EGGERS   *Harvard University*
JENS HAINMUELLER   *Massachusetts Institute of Technology*

M
any recent studies show that firms profit from connections to influential politicians, but less is known about how much politicians financially benefit from wielding political influence. We estimate the returns to serving in Parliament, using original data on the estates of recently deceased British politicians. Applying both matching and a regression discontinuity design to compare Members of Parliament (MPs) with parliamentary candidates who narrowly lost, we find that serving in office almost doubled the wealth of Conservative MPs, but had no discernible financial benefits for Labour MPs. Conservative MPs profited from office largely through lucrative outside employment they acquired as a result of their political positions; we show that gaining a seat in Parliament more than tripled the probability that a Conservative politician would later serve as a director of a publicly traded firm—enough to account for a sizable portion of the wealth differential. We suggest that Labour MPs did not profit from office largely because trade unions collectively exerted sufficient control over the party and its MPs to prevent members from selling their services to other clients.

Exercise 2: Assessing Discontinuities

Conservative Candidates

Exercise 2: Assessing Discontinuities

Please answer the following questions:

1. The authors use a 15 percentage point bandwidth around the cutoff point. What are the costs and benefits for wider vs. narrower bandwidths? Do you find this bandwidth credible?

2. Causal inference in observational settings often relies on "as-if" random assignment in order to address OVB. Is it necessary to assume that who wins and who loses among all candidates within the bandwidth is essentially random? Why or why not?

3. What do you see as the most likely threat to the validity of a regression discontinuity design in this context?

Exercise 2: Assessing Discontinuities

1. Wider bandwidths allow us to use more data to estimate the effect, shrinking our standard errors. However, using a wide bandwidth makes individuals above and below the threshold less comparable, introducing bias to the causal effect. A 15pp bandwidth may be larger than is appropriate because individuals who lost an election by 15 points are likely substantially different from those who are on the cusp of winning/losing.

2. Only the victory/loss of individuals just above and below the threshold needs to be essentially random, not all individuals in the bandwidth.

3. If candidates just above the threshold are differentially able to use their resources to narrowly win election compared to candidates just below, then candidates above/below the cutoff are no longer comparable.

Estimating and Plotting RD in R

Now I'll go through RD estimation and plotting in R using `feols`. Load the simulated data file: `rd.dta`.
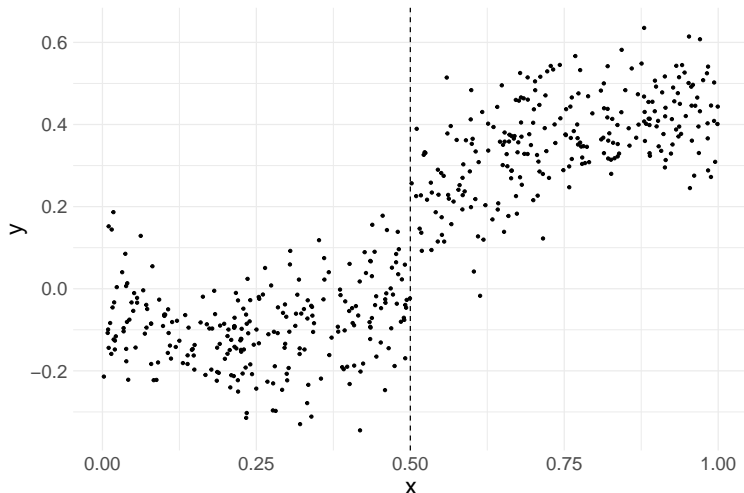
```
# ------------------------- #
# RD Estimation & Plotting
# ------------------------- #
# Load data
rd_data <- read_dta("rd.dta")
```

Variables

- x: running variable
- treat: treatment indicator
- y: outcome

Estimating and Plotting RD in R

Let's start by taking a look at the data, specifically the running variable and the outcome.

Estimating and Plotting RD in R

```
# Recenter cutoff at zero
rd_data <- mutate(rd_data, x_cent = x - 0.5)

# Estimate interacted regression
fit_rd <- feols(y ~ x_cent + treat + x_cent:treat,
                rd_data,
                vcov = "hetero")
fit_rd
```

```
OLS estimation, Dep. Var.: y
Observations: 500
Standard-errors: Heteroskedasticity-robust
             Estimate Std. Error  t value   Pr(>|t|)
(Intercept) -0.065664   0.013262 -4.95127 1.0124e-06 ***
x_cent       0.098011   0.045610  2.14892 3.2124e-02 *
treat        0.316241   0.018993 16.65028  < 2.2e-16 ***
x_cent:treat 0.338082   0.062893  5.37550 1.1777e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.096975   Adj. R2: 0.848589
```
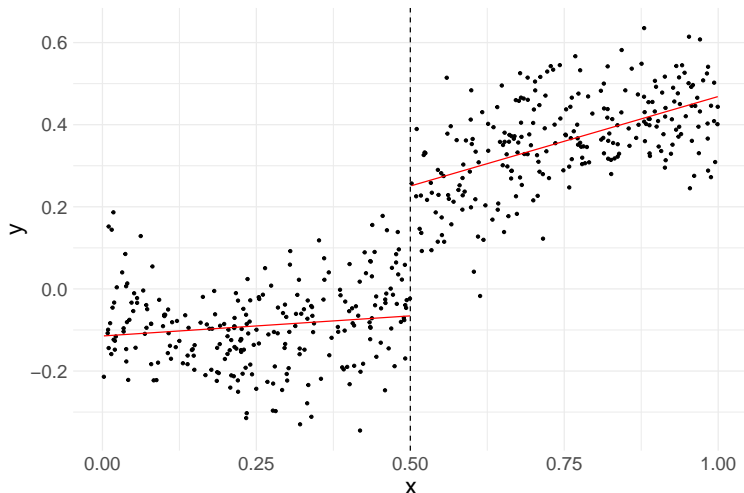
Estimating and Plotting RD in R

```
# Get predicted values
predictions <- fit_rd$fitted.values
```

Estimating and Plotting RD in R

Our RD estimates indicate a local average treatment effect of 0.316, but the true treatment effect is 0.2. (I only know this because I simulated the data.)

Who is responsible for this bias? Mr. Polynomial.

The fact that we didn't account for the curvature of the underlying relationship can bias the coefficient on `treat`. We can deal with this bias by allowing for a more flexible relationship between the running variable and the dependent variable.

Estimating and Plotting RD in R

```
# Estimate interacted regression
fit_rd_poly <- feols(y ~ treat +
                     x_cent * treat +
                     x_cent^2 * treat +
                     x_cent^3 * treat,
                   rd_data,
                   vcov = "hetero")

broom::tidy(fit_rd_poly) %>% filter(term == "treat")

# A tibble: 1 x 5
  term  estimate std.error statistic       p.value
  <chr>    <dbl>     <dbl>     <dbl>         <dbl>
1 treat    0.204    0.0344      5.92 0.00000000606
```
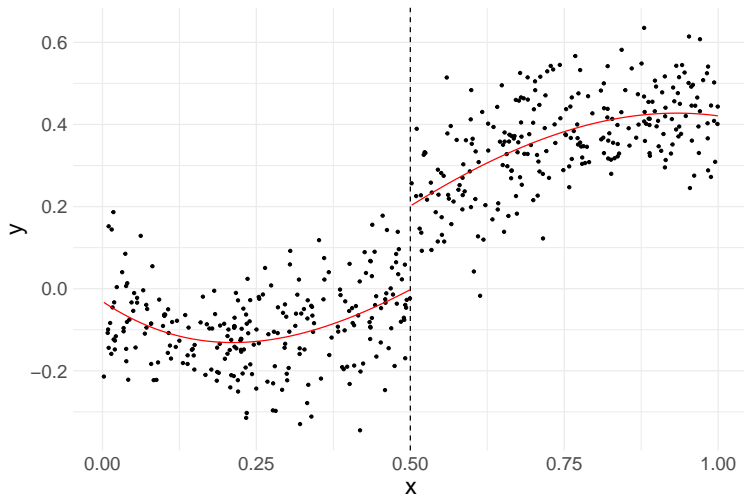
Estimating and Plotting RD in R

We can also get fitted values from this non-linear regression and plot them.

Estimating and Plotting RD in R

Lastly, we can try using a smaller bandwidth for estimation.

```
# Keep observations in bandwidth (0.15 points)
rd_data_bw15 <- filter(rd_data, abs(x - 0.5) <= 0.15)

# Estimate interacted regression
fit_rd_bw15 <- feols(y ~ x_cent + treat + x_cent:treat,
                      rd_data_bw15,
                      vcov = "hetero")

broom::tidy(fit_rd_bw15) %>% filter(term == "treat")
```

```
# A tibble: 1 x 5
  term  estimate std.error statistic  p.value
  <chr>    <dbl>     <dbl>     <dbl>    <dbl>
1 treat    0.229    0.0339      6.76 3.85e-10
```

Estimating and Plotting RD in R