

Causal Inference Workshop

Week 3 - Instrumental Variables and Regression Discontinuity

Causal Inference Workshop

February 6, 2024

Fanyu Wang, fw397@columbia.edu - SDEV 9280

Workshop outline

A. Causal inference fundamentals

- Modeling assumptions matter too
- Conceptual framework (potential outcomes framework)

B. Design stage: common identification strategies

- IV + RDD [coding]
- DiD, DiDiD, Event Studies, New TWFE Lit [coding]
- Synthetic Control / Synthetic DiD [coding]

C. Analysis stage: strengthening inferences

- Limitations of identification strategies, pre-estimation steps
- Estimation [controls] and post-estimation steps [supporting assumptions]

D. Other topics in causal inference and sustainable development

- Inference (randomization inference, bootstrapping)
- Weather data regressions, other common/fun SDev topics [coding]
- Remote sensing data, other common/fun SDev topics

Causal inference roadmap

- *Potential outcomes* [framework] [last week]
 - Causal effect is the difference between two potential outcomes
 - We can't observe this difference, but can see differences in average observed outcomes
 - If **(conditional) independence assumption** holds, can estimate unbiased ATT
- *Identification* [application/implementation] [today]
 - In most empirical settings, IA and CIA do not hold, which is why we need an **identification strategy**
 - Want to eliminate selection bias (identification problem)
- *Estimation* [application/implementation]
 - (Usually) use linear regression model
 - $\hat{\beta}_{OLS}$ unbiased estimator for ATT if e is uncorrelated with treatment (regression problem)

Outline

Workshop outline

Canonical identification strategies

Instrumental variables

Regression discontinuity

Hierarchy of common identification methods

Most common identification methods:

- **Randomized experiments (RCT)** - natural randomization of treatment D
- **Instrumental variables (IV) or regression discontinuity (RD)** - instrument or discontinuity that induces exogenous variation in treatment status
- **Difference-in-differences (DiD), event studies, synthetic control methods (SCM)** - research designs that assume or construct parallel trends
- **Matching estimators** - strategies solely based on matching are much less credible, but matching can complement natural or quasi-experimental design

Hierarchy of common identification methods

For each, we will review:

- Assumed data generating process (DGP)
- Identifying assumptions
- Estimand (treatment of interest)
- Estimator used
- Canonical examples
- Best practices
- Strengths and weaknesses
- *SDev-y examples*
- *Coding implementation / exercises*

→ relationship between actual observed outcomes (Y_i) and the conceptual potential outcomes (Y_i^0, Y_i^1), e.g. why is our estimation able to recover a *causal* treatment effect?

Outline

Workshop outline

Canonical identification strategies

Instrumental variables

Regression discontinuity

Outline

Workshop outline

Canonical identification strategies

Instrumental variables

Homogeneous Effects

Heterogeneous Effects (Estimating LATE)

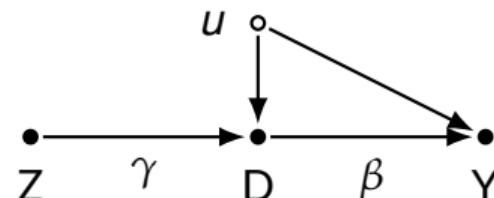
Best Practice, Strengths, and Weakness

Regression discontinuity

Instrumental variables, DGP under Homogeneous Effects

$$Y_i = \alpha + \beta D_i + u_i, \quad \text{cov}[D_i, u_i] \neq 0$$

- D_i is endogenous
- But, there exists a binary instrument Z_i that is a random source of variation in D_i , it "assigns" or changes the probability of treatment
 - We use the instrument to isolate variation in D that is unrelated to u and recover β
- Backdoor path between D and Y
(open, selection on unobservables)
- But mediating path from Z to Y
(Z affects Y "only through" D)



Identification assumptions

- Identifying assumptions

A1. Exclusion Restriction	$\text{cov}[Z_i, u_i] = 0$
A2. Relevance	$\text{cov}[Z_i, D_i] \neq 0$

- Projection Model: $D_i = \delta + \gamma Z_i + v_i$, $\text{cov}[Z_i, v_i] = 0$.
- Because $\text{cov}[Z_i, v_i] = 0$ and A1, $\text{cov}[Z_i, \varepsilon_i] = 0$, we have $\beta\gamma$ identified from below

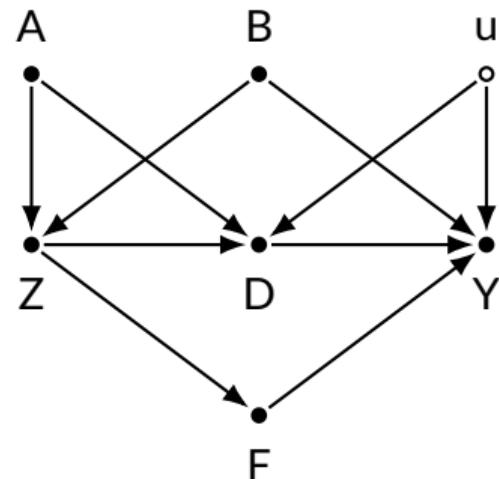
$$\begin{aligned} Y_i &= \alpha + \beta(\delta + \gamma Z_i + v_i) + u_i \\ &= (\alpha + \beta\delta) + \beta\gamma Z_i + \underbrace{\beta v_i + u_i}_{\varepsilon_i} \end{aligned}$$

- A2 guarantees that $\gamma \neq 0$

Instrumental variables, more on assumptions

- **Relevance** - show F-statistic
- **Validity / exclusion restriction¹** (Z affects Y only through D) - trickier! why?

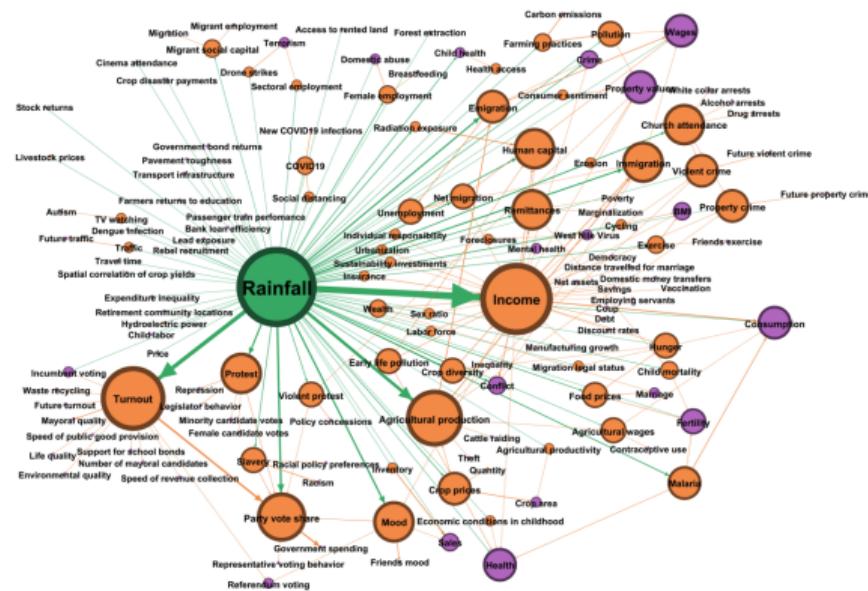
- Problem is unobserved u , have instrument Z , but...
- We want all open paths from Z to Y to contain D
- If we don't control for A , that's okay
- What about B and F ?
 - B some confounder, control for it
 - F issue too, because it means variation in D driven by Z is closely related to F too → mixing together effect of D and F



1. called this because Z can be excluded after $Z \rightarrow D$ path included

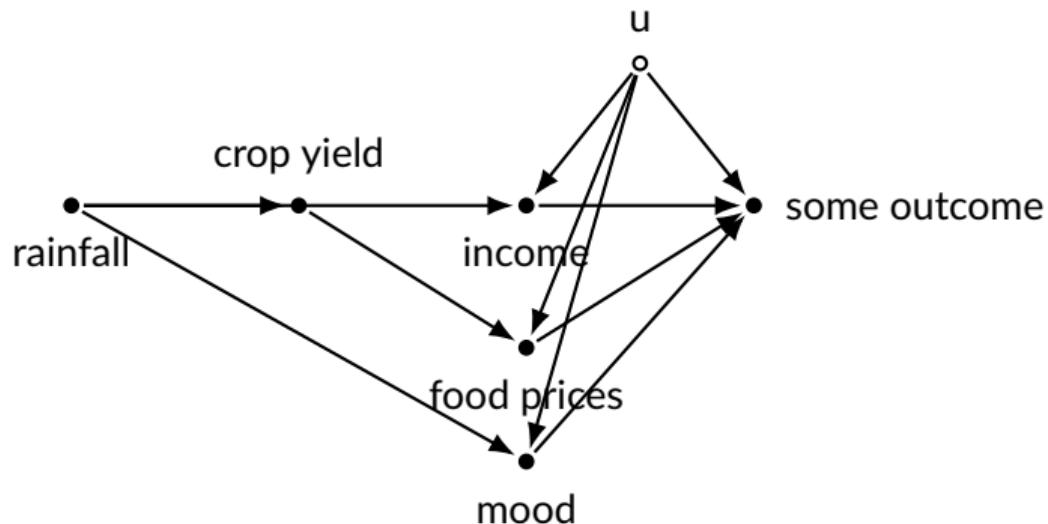
Instrumental variables, more on assumptions

- **Validity / exclusion restriction** (Z affects Y only through D) - trickier! why?
 - Validity of instruments can be a big concern
 - Plenty of instruments that later turn out to be not so great...
 - For example, **rainfall** ([Mellon 2023](#))



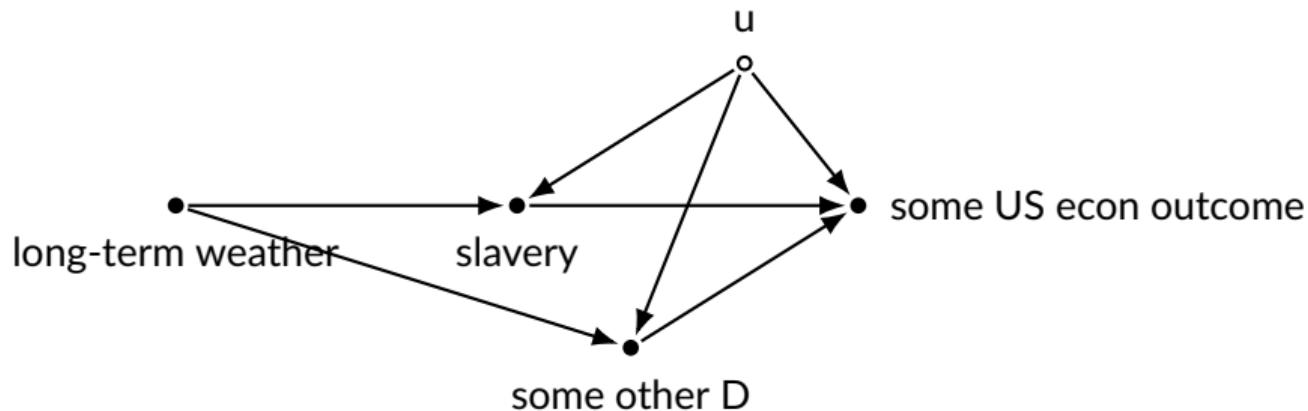
Instrumental variables, more on assumptions

- **Validity / exclusion restriction** (Z affects Y only through D) - trickier! why?
 - Validity of instruments can be a big concern
 - Plenty of instruments that later turn out to be not so great...
 - For example, **rainfall** ([Mellon 2023](#))



Instrumental variables, more on assumptions

- **Validity / exclusion restriction** (Z affects Y only through D) - trickier! why?
 - Validity of instruments can be a big concern
 - Plenty of instruments that later turn out to be not so great...
 - For example, **rainfall** ([Mellon 2023](#))



Instrumental variables, estimand and estimator

- β is identified under the two assumptions

$$\beta = \frac{\text{cov}[Y_i, Z_i]}{\text{cov}[D_i, Z_i]}$$

- Estimator
 - Sample analog called Wald estimator, $\hat{\beta}_W = \frac{\hat{\text{cov}}[Y_i, Z_i]}{\hat{\text{cov}}[D_i, Z_i]}$
 - Numerically equivalent to two-stage least squares (2SLS) estimator $\hat{\beta}_{2SLS}$ obtained through

$$\text{1st stage: } D_i = \delta + \gamma Z_i + v_i \rightarrow \hat{D}_i = \hat{\mathbb{E}}[D_i | Z_i]$$

$$\text{2nd stage: } Y_i = \tilde{\alpha} + \tilde{\beta} \hat{D}_i + e_i$$

- Note: SEs of the 2nd stage wouldn't give correct SEs (need to adjust for two stages of estimation); 2SLS packages do adjustment automatically, so use those or bootstrap

Instrumental variables, canonical examples

- Judge harshness as an instrument for punishment (Aizer and Doyle 2015, QJE)
→ Juvenile incarceration → substantially lower high school completion rates & higher adult incarceration rates



Johanna Rickne
@johannarickne

...

Banning the purchase of sex 🚫 DOES NOT 🚫 increase cases of reported rape.

A re-analysis of Ciacci (2024) shows that the paper's headline result comes from an erroneous use of Stata's regression command.

A thread from @Jopieboy, @OlleFolke, and me 1/11



John B. Holbein @JohnHolbein1 · Mar 17

Banning the purchase of sex increases cases of rape.

link.springer.com/article/10.1007/s10640-024-00944-w...

evidence from Sweden

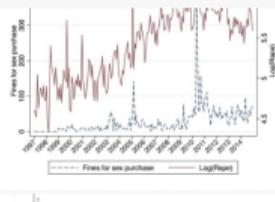
Riccardo Cicconi

Received: 3 May 2023 / Accepted: 5 January 2024

© The Author(s). Under exclusive license to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

This paper leverages the lifting of a ban on the purchase of sex to assess its impact on rape offenses. Relying on Swedish high frequency data from 1997 to 2016, I find that the ban increases the number of rapes by around 44–45%. The results are robust to several model specifications and different identification strategies. The increase reflects a boost in completed rapes both in the short- and long-run. However, it is not accompanied by a decrease in the number of pings. Taken together, the evidence suggests that the ban increased the number of rapes not through the supply of prostitution but rather changes in the demand for prostitution due to the ban. The results here have the opposite sign but larger magnitudes in absolute value than results in the literature on the decriminalization of prostitution.



Johanna Rickne
@johannarickne

...

Most of Stata's regression commands drop the treatment variable from

Outline

Workshop outline

Canonical identification strategies

Instrumental variables

Homogeneous Effects

Heterogeneous Effects (Estimating LATE)

Best Practice, Strengths, and Weakness

Regression discontinuity

Instrumental variables, potential outcomes, allowing heterogeneous Effects

- $D_i = \delta + \gamma_i Z_i + v_i$, $Y_i = \alpha + \beta_i D_i + u_i$
- Treatment assignment ($Z_i \in \{0, 1\}$) and treatment realization ($D_i \in \{0, 1\}$) - how does instrument affect treatment status?
 - Compliers: Treatment status affected by instrument in the correct direction
 $\rightarrow D_i^1 = 1; D_i^0 = 0$
 - Defiers: Treatment status affected by instrument in the wrong direction
 $\rightarrow D_i^1 = 0; D_i^0 = 1$
 - Never-takers: Never take treatment, treatment status not affected by instrument
 $\rightarrow D_i^1 = 0; D_i^0 = 0$
 - Always-takers: Always take treatment, treatment status not affected by instrument
 $\rightarrow D_i^1 = 1; D_i^0 = 1$
- Researcher can only observe Z_i and D_i , not these groups

Identification assumptions for LATE

- Identifying assumptions

A1. independence (of Z)	$Z_i \perp\!\!\!\perp (D_i^0, D_i^1, Y_i^0, Y_i^1)$	This gives $\text{cov}[Z_i, v_i] = 0, \text{cov}[Z_i, u_i] = 0$
A2. relevance (of Z)	$\text{cov}[Z_i, D_i] \neq 0$	Z does affect D ; equivalent to $P(D_i = 1 Z_i = 1) \neq P(D_i = 1 Z_i = 0)$
A3. monotonicity (of Z on D)	no defiers	Z is an incentive, does not discourage treatment; equivalent to $D_i^1 \geq D_i^0$

- A1 is not implied by random assignment of Z . Imaging four potential outcomes, $\{Y_i^{(z,d)} : z = 0, 1, d = 0, 1\}$. We need to assume

$$Z_i \perp\!\!\!\perp (Y_i^{(0,0)}, Y_i^{(0,1)}, Y_i^{(1,0)}, Y_i^{(1,1)}, D_i^0, D_i^1)$$

and

$Y_i^{(z,d)} = Y_i^{(z',d)} = Y_i^d$ for all z, z', d (sometime separately called Exclusion restriction).

LATE: Compliance Types by Treatment and Instrument

Table: With A3: Monotonicity Assumption

	Z=0	Z=1
D=0	compiler or never-taker	never-taker
D=1	always-taker	compiler or always=taker

Let π_c , π_n , and π_a be population proportions of compliers, never-takers, and always takers. Then

$$\begin{aligned}\pi_a &= \mathbb{E}[D|Z = 0] \\ \pi_a + \pi_c &= \mathbb{E}[D|Z = 1] \\ \pi_n &= 1 - \mathbb{E}[D|Z = 1].\end{aligned}$$

The first two equations give $\pi_c = \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]$. By A2 and A3, $\pi_c > 0$.

LATE: Identification

Because $D = (1 - Z)D^0 + ZD^1 = D^0 + Z(D^1 - D^0)$,

$$Y = (1 - D)Y^0 + DY^1 = Y^0 + D^0(Y^1 - Y^0) + Z(D^1 - D^0)(Y^1 - Y^0).$$

Then by A1,

$$\begin{aligned}\mathbb{E}[Y|Z] &= \mathbb{E}[Y^0|Z] + \mathbb{E}[D^0(Y^1 - Y^0)|Z] + \mathbb{E}[Z(D^1 - D^0)(Y^1 - Y^0)|Z] \\ &= \mathbb{E}[Y^0] + \mathbb{E}[D^0(Y^1 - Y^0)] + Z\mathbb{E}[(D^1 - D^0)(Y^1 - Y^0)].\end{aligned}$$

Thus, $\mathbb{E}[Y|Z = 1] - \mathbb{E}[Y|Z = 0] = \mathbb{E}[(D^1 - D^0)(Y^1 - Y^0)]$

LATE: Identification (Cont'd)

$$\begin{aligned} & \mathbb{E}[(D^1 - D^0)(Y^1 - Y^0)] \\ &= 1 \cdot \mathbb{E}[Y^1 - Y^0 | D^1 - D^0 = 1] \mathbb{P}(D^1 - D^0 = 1) \\ &\quad + 0 \cdot \mathbb{E}[Y^1 - Y^0 | D^1 - D^0 = 0] \mathbb{P}(D^1 - D^0 = 0) \\ &\quad + (-1) \cdot \mathbb{E}[Y^1 - Y^0 | D^1 - D^0 = -1] \mathbb{P}(D^1 - D^0 = -1) \\ &= \mathbb{E}[Y^1 - Y^0 | D^1 - D^0 = 1] \mathbb{P}(D^1 - D^0 = 1) \end{aligned}$$

because of A3. Note that $\mathbb{P}(D^1 - D^0 = 1) = \pi_c = \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0] > 0$, thus

$$\mathbb{E}[Y^1 - Y^0 | D^1 - D^0 = 1] = \frac{\mathbb{E}[Y|Z = 1] - \mathbb{E}[Y|Z = 0]}{\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]}.$$

Outline

Workshop outline

Canonical identification strategies

Instrumental variables

Homogeneous Effects

Heterogeneous Effects (Estimating LATE)

Best Practice, Strengths, and Weakness

Regression discontinuity

Instrumental variables, best practices, strengths and weaknesses

- Best practices
 - Support **relevance** assumption by showing a large F-statistic for the 1st stage ($F > 10$, but bigger is better, bigger F = “stronger” instrument) see Stock and Yogo (2002) for more!
 - In case of weak IV (if you don’t want to give up), try approach more robust to weak instruments (see Andrews et al. 2019)
 - As in any observational study, adjust for all other *relevant* pre-treatment variables (predictors of Y not affected by D), include the same variables in both stages
 - Different valid instruments select different set of compliers, leading to different estimands and estimates; think of group of compliers selected and make sure instrument is relevant w.r.t. policy of interest
 - For models non-linear in D , properties of 2SLS do not necessarily hold, may want to consider alternative estimation strategies (e.g., control function method)
- Strengths & weaknesses

Instrumental variables, best practices, strengths and weaknesses

- Best practices
- Strengths & weaknesses
 - + Compelling identification strategy
 - $\hat{\beta}_{IV}$ less efficient than OLS, precision further decreases with weak instruments
 - $\hat{\beta}_{IV}$ has “finite sample bias”, which stems from randomness in estimates of \hat{D}_i and increases with weakness and number of instruments
 - Weak instruments can render $\hat{\beta}_{IV}$ considerably less efficient and even more biased than $\hat{\beta}_{OLS}$ (Andrews et al. 2019)
 - In many settings (e.g., non-linear D), 2SLS can be very biased
 - + Can use IVs to address attenuation bias that may result from measurement error in D (e.g., Krueger and Lindahl 2001)

Outline

Workshop outline

Canonical identification strategies

Instrumental variables

Regression discontinuity

(Sharp) Regression discontinuity, DGP

$$Y_i = \alpha + \beta D_i + f(X_i, \phi) + u_i$$

- Treatment D_i is not randomly assigned, it is deterministic, but *discontinuous* along a continuous pretreatment **running variable** X_i a **cutoff** c (e.g., $D_i = \mathbb{1}\{X_i \geq c\}$)
- D_i deterministic function of X_i (no value of X_i with both treatment and control).
- We only observe the outcome under control, $Y_i(0)$, for those units whose running variable (also called **score**) is below the cutoff, and we only observe the outcome under treatment, $Y_i(1)$, for those units whose score is above the cutoff.
- Look at data only in a small neighborhood around c (cutoff), the **bandwidth**

(Sharp) Regression discontinuity, potential outcomes

- The following statement shows up often, but it is actually not part of the identification assumptions for the typical RD with continuous X , it is for another identification strategy called Local Randomization, which could also be categorized as RD, but typically with discrete X (see Cattaneo et al. (2024)).
 - Average outcome of those right below the cutoff (who are denied treatment) are compared to those right above the cutoff (who receive the treatment) (i.e., $\mathbb{E}[Y_i(d)|X_i < c] = \mathbb{E}[Y_i(d)|X_i \geq c]$ for $d = 0, 1$)
- Real assumption needed: $\mathbb{E}[Y_i(1)|X_i]$ and $\mathbb{E}[Y_i(0)|X_i]$ continuous in X_i at c .

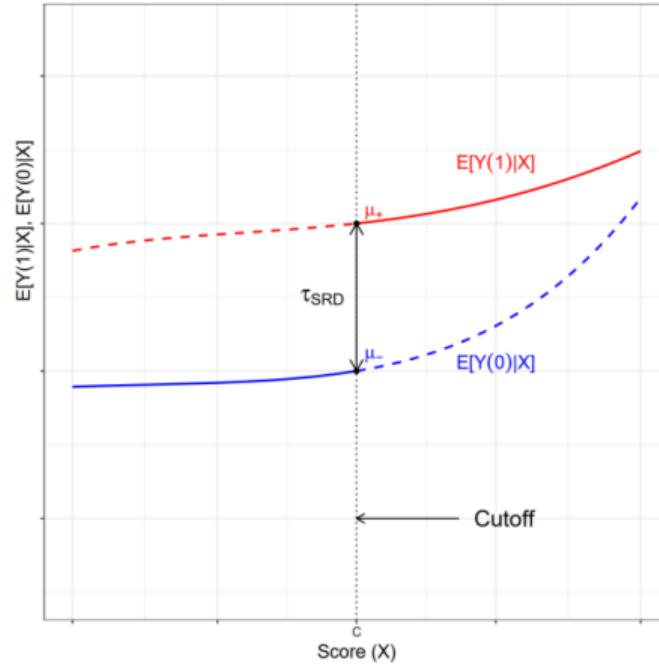


Figure: RD Treatment Effect in Sharp RD Design
(Source: Cattaneo et al. (2020))

(Sharp) Regression discontinuity, identifying assumptions

- Identifying assumptions

A1. local continuity	$\mathbb{E}[Y_i^1 X_i]$ and $\mathbb{E}[Y_i^0 X_i]$ continuous in X_i at c	other determinants of Y don't jump at c
A2. relevance	$D_i = \mathbb{1}[X_i \geq c]$	discontinuity in the dependence of D_i on X_i

→ We can attribute a jump in Y_i at c to the causal effect of D_i

Regression discontinuity, canonical examples

- Explicit cutoffs in programs (e.g., income in means-tested programs, test scores in gifted-and-talented programs)
- Geographic cutoffs (e.g., school-zone boundaries, such as Black (1999), time zone borders, etc.)
 - e.g., Black (1999) uses house values near elementary school zone boundaries and finds parents are willing to pay 2.5% more for 5% increase in school test scores
- Election cutoffs (e.g., need 50% for win)

(Sharp) Regression discontinuity, estimand and estimator

- Estimand

$$\beta_{RD} = \lim_{x \rightarrow c^+} \mathbb{E}[Y_i | X_i = x] - \lim_{x \rightarrow c^-} \mathbb{E}[Y_i | X_i = x] = \dots = \mathbb{E}[Y_i^1 - Y_i^0 | X_i = c]$$

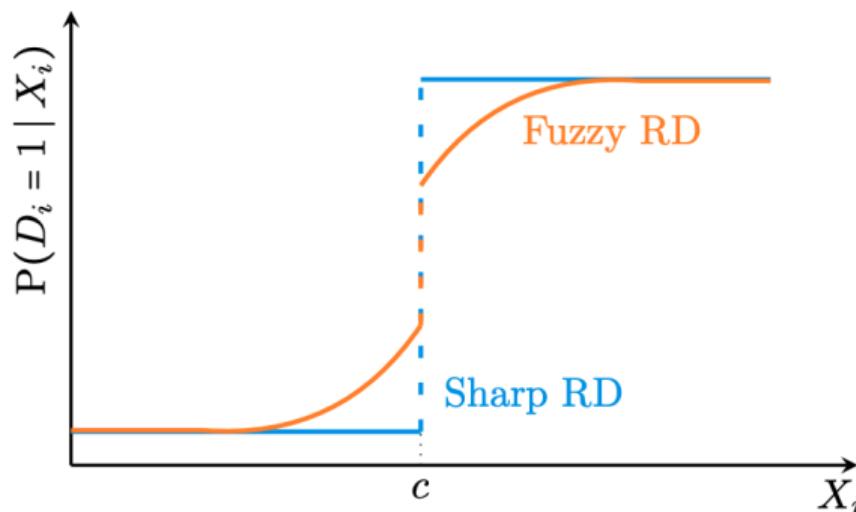
- Estimator

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

- Use flexible functional forms for $f(X_i)$, such as:
 - local linear regression model: $Y_i = \alpha + \beta D_i + \gamma_1(X - c) + \gamma_2(X - c)D + e_i$, with $c - h \leq X \leq c + h$
 - polynomial regression model with low-degree polynomial (e.g., quadratic, as higher order polynomials can lead to overfitting and introduce bias, see Gelman and Imbens 2019)

(Fuzzy) Regression discontinuity, estimand and estimator

- In a fuzzy RD, there is imperfect compliance, and at $X_i \geq c$, there is a jump but not in treatment assignment but in the *probability* of treatment assignment ($P(D_i = 1|X)$)
→ Discontinuity becomes an instrumental variable for the treatment status D_i



(a) RD treatment assignment (sharp & fuzzy)

(Fuzzy) Regression discontinuity, estimand and estimator

- In a fuzzy RD, there is imperfect compliance, and at $X_i \geq c$, there is a jump but not in treatment assignment but in the *probability* of treatment assignment ($P(D_i = 1|X)$)
→ Discontinuity becomes an instrumental variable for the treatment status D_i
- Estimand

$$\beta_{RD} = \lim_{x \rightarrow c^+} \mathbb{E}[Y_i | X_i = x] - \lim_{x \rightarrow c^-} \mathbb{E}[Y_i | X_i = x] = \dots = \mathbb{E}[Y_i^1 - Y_i^0 | X_i = c]$$

- Estimator (estimate using 2SLS)

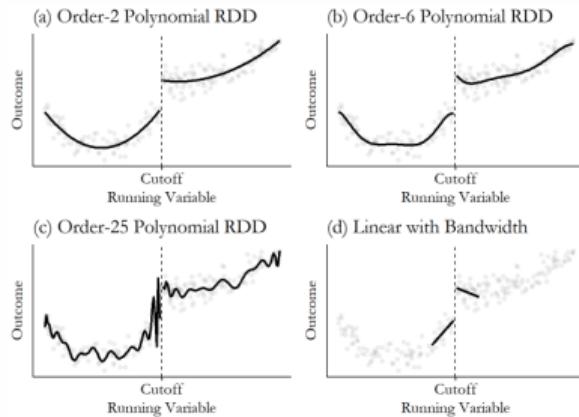
$$\text{1st stage: } D_i = \delta + \gamma Z_i + f(X_i) + u_i \rightarrow \hat{D}_i = \hat{\mathbb{E}}[D_i | X_i]$$

$$\text{2nd stage: } Y_i = \tilde{\alpha} + \tilde{\beta} \hat{D}_i + f(X_i) + e_i$$

Regression discontinuity, best practices, strengths and weaknesses

- Best practices

- Choice of $f()$: $f()$ is unknown, so misspecification of the functional form of the DGP may bias the estimator, do robustness checks



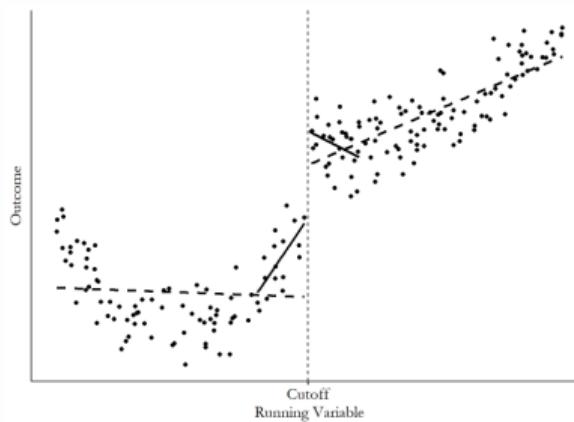
Source: <https://theeffectbook.net>

- Bandwidth choice can also influence estimate, do robustness checks
- As in any observational study, adjust for all relevant pre-treatment variables

Regression discontinuity, best practices, strengths and weaknesses

- Best practices

- Choice of $f()$: $f()$ is unknown, so misspecification of the functional form of the DGP may bias the estimator, do robustness checks
- Bandwidth choice can also influence estimate, do robustness checks



Source: <https://theeffectbook.net>

- As in any observational study, adjust for all relevant pre-treatment variables

Regression discontinuity, best practices, strengths and weaknesses

- Best practices
 - Choice of $f()$: $f()$ is unknown, so misspecification of the functional form of the DGP may bias the estimator, do robustness checks
 - Bandwidth choice can also influence estimate, do robustness checks
 - As in any observational study, adjust for all relevant pre-treatment variables
- Strengths & weaknesses
 - + Similar to a local randomized experiment and thereby require weak assumptions
 - + All about finding “jumps” in the probability of treatment as we move along some X ; much potential in economic applications as geographic boundaries and administrative or organizational rules often create usable discontinuities
 - Risk being underpowered
 - Parameter estimates are very “local”, so their external validity may be low

Questions? Comments?

Thank you!

References I

Heavily based on Claire Palandri's 2022 version and Anna Papp's 2024 version of the Causal Inference Workshop.

- Aizer, Anna, and Joseph J. Doyle. 2015. "JUVENILE INCARCERATION, HUMAN CAPITAL, AND FUTURE CRIME: EVIDENCE FROM RANDOMLY ASSIGNED JUDGES." *The Quarterly Journal of Economics* 130 (2): 759–804. ISSN: 00335533, 15314650, accessed January 26, 2024. <https://www.jstor.org/stable/26372613>.
- Andrews, Isaiah, James H. Stock, and Liyang Sun. 2019. "Weak Instruments in Instrumental Variables Regression: Theory and Practice." *Annual Review of Economics* 11 (1): 727–753. <https://doi.org/10.1146/annurev-economics-080218-025643>. eprint: <https://doi.org/10.1146/annurev-economics-080218-025643>.
- Angrist, Joshua D. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." *The American Economic Review* 80 (3): 313–336. ISSN: 00028282, accessed January 26, 2024. <http://www.jstor.org/stable/2006669>.
- Angrist, Joshua D., and Alan B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics* 106 (4): 979–1014. ISSN: 00335533, 15314650, accessed January 26, 2024. <http://www.jstor.org/stable/2937954>.
- Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *The Quarterly Journal of Economics* 114 (2): 577–599. ISSN: 00335533, 15314650, accessed January 26, 2024. <http://www.jstor.org/stable/2587017>.
- Cattaneo, Matias D., Nicolas Idrobo, and Rocío Titiunik. 2024. *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press.

References II

- Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2020. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press.
- Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics* 37 (3): 447–456. <https://doi.org/10.1080/07350015.2017.1366909>. eprint: <https://doi.org/10.1080/07350015.2017.1366909>.
- Krueger, Alan B., and Mikael Lindahl. 2001. "Education for Growth: Why and for Whom?" *Journal of Economic Literature* 39 (4): 1101–1136. <https://doi.org/10.1257/jel.39.4.1101>. <https://www.aeaweb.org/articles?id=10.1257/jel.39.4.1101>.
- Stock, James H, and Motohiro Yogo. 2002. *Testing for Weak Instruments in Linear IV Regression*. Working Paper, Technical Working Paper Series 284. National Bureau of Economic Research. <https://doi.org/10.3386/t0284>. <http://www.nber.org/papers/t0284>.