Causal Inference Workshop

Week 3 - Instrumental Variables and Regression Discontinuity

Causal Inference Workshop

February 2, 2024

Anna Papp, ap3907@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

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

Instrumental variables, DGP

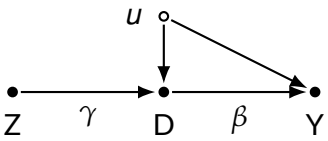
$$Y_i = \alpha + \beta D_i + u_i, \quad cov[D_i, u_i] \neq 0$$

- 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 e and recover β

$$D_i = \delta + \gamma Z_i + v_i$$
$$Y_i = \alpha + \beta D_i + u_i, \quad cov[D_i, u_i] \neq 0$$

- Backdoor path between D and Y (open, selection on unobservables)
- But mediating path from Z to Y (Z affects Y "only through" D)



Instrumental variables, potential outcomes

- 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

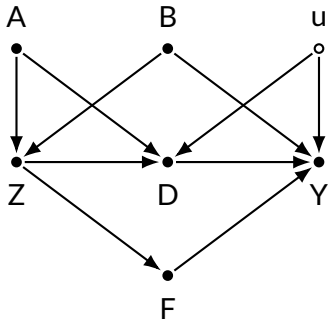
Instrumental variables, identifying assumptions

- Identifying assumptions

A1. independence (of Z)	$cov[Z_i, v_i] = 0$	no unmeasured confounder affecting both instrument & outcome
A2. exclusion restriction	$cov[Z_i, u_i] = 0$	no direct effect of Z on Y; Z affects Y only through D
A3. relevance (of Z)	$cov[Z_i, D_i] \neq 0$	Z does affect D
A4. monotonicity (of Z on D)	no defiers	Z is an incentive, does not discourage treatment

Instrumental variables, more on assumptions

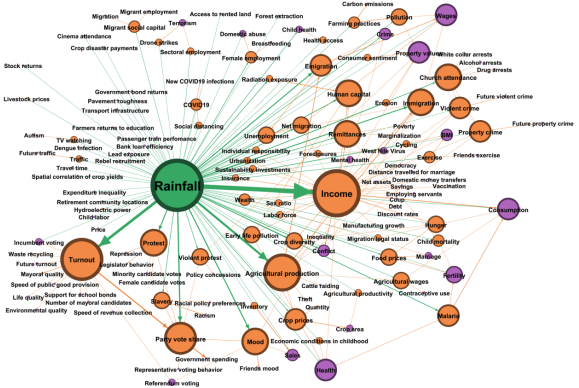
- **Relevance** - show F-statistic
- **Validity / exclusion restriction**[1] (Z affects Y only through D) - trickier! why?
  - Problem is unobserved u, have instrument Z, but...
  - We want all open paths from Z to Y to contain D
  - If we don't control for A, that's okay
  - What about B and F?
    - 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

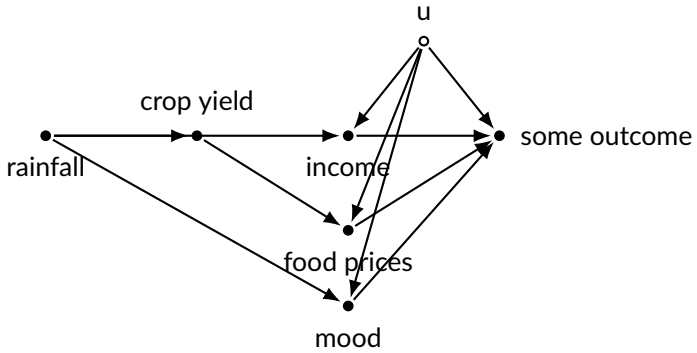
Instrumental variables, more on assumptions

- **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)



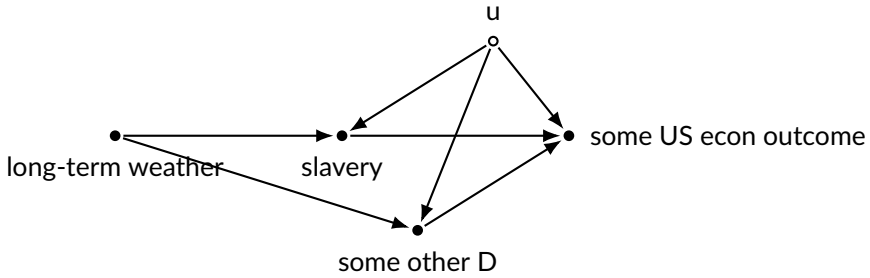
Instrumental variables, more on assumptions

- **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, 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

- Military drafts as an instrument for military service (Angrist 1990, AER)
  → Earnings of white veterans 15% less than nonveterans, even long after service

- Compulsory schooling as an instrument for years of education (Angrist and Krueger 1991, QJE)
  → IV estimate of return to education close to OLS estimate in this case

- "Bartik Shift-Share" instrument (Bartik 1991)
  Goldsmith-Pinkham et al. paper: https://paulgp.github.io/papers/bartik_gpss.pdf

Instrumental variables, estimand and estimator

- Estimand

$$\beta_{IV} = \frac{cov[Y_i, Z_i]}{cov[D_i, Z_i]} = \ldots = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \ldots$$
$$= \underbrace{\mathbb{E}[Y_i^1 - Y_i^0 | D_i^0 = 0, D_i^1 = 1]}_{\text{LATE on the compliers}}$$

- Estimator
  - Sample analog called Wald estimator, $\hat{\beta}_W = \frac{\hat{cov}[Y_i, Z_i]}{\hat{cov}[D_i, Z_i]}$
  - 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, 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

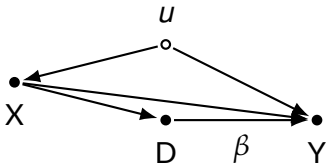
- Strengths & weaknesses
  - \+ Compelling identification strategy
  - \- $\hat{\beta}_{IV}$ less efficient than OLS, precision further decreases with weak instruments
  - \- $\hat{\beta}_{IV}$ has "finite sample bias", which stems from randomness in estimates of $\hat{D}_i$ and increases with weakness 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

(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, such that there is "local randomization" around a cutoff c (e.g., $D_i = \mathbb{1}\{X_i \geqslant c\}$)
- D_i deterministic function of X_i (no value of X_i with both treatment and control), so must extrapolate across X_i
- Look at data only in a small neighborhood around c (cutoff), the bandwidth



(Sharp) Regression discontinuity, potential outcomes

- Average outcome of those right below the cutoff (who are denied treatment) are compared to those right above the cutoff (who receive the treatment)

(Sharp) Regression discontinuity, identifying assumptions

- Identifying assumptions

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

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

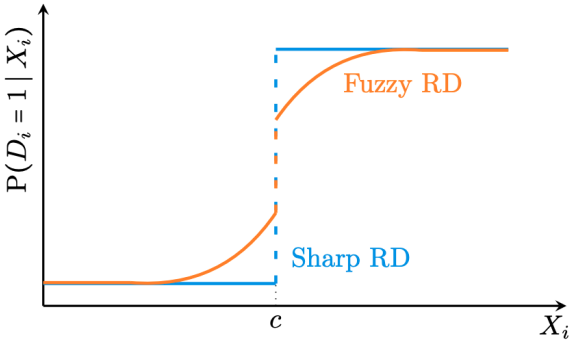
- Estimator

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

  - 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 $c - h \leqslant X \leqslant c + h$
    - 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 $X_i \geqslant c$, 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 $X_i \geqslant c$, there is a jump but not in treatment assignment but in the *probability* of treatment assignment ($P(D_i = 1|X)$)
  $\rightarrow$ Discontinuity becomes an instrumental variable for the treatment status D_i

- Estimand

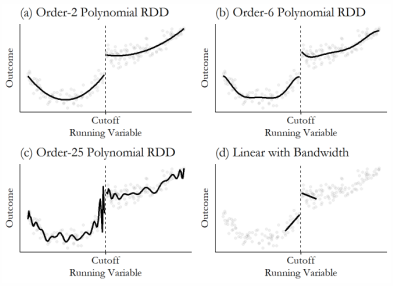
$$\beta_{RD} = \lim_{x \to c^+} \mathbb{E}[Y_i|X_i = x] - \lim_{x \to c^-} \mathbb{E}[Y_i|X_i = x] = ... = \mathbb{E}[Y_i^1 - Y_i^0|X_i = c]$$

- 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

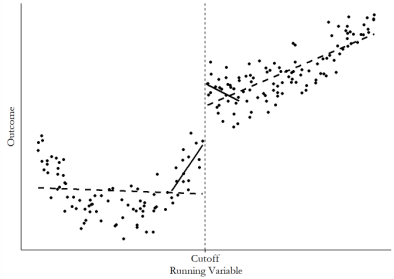


Source: https://theeffectbook.net

  - Bandwith 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
  - Bandwith 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
  - Bandwith 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 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. 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.

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. https://doi.org/10.1080/07350015.2017.1366909.

References II

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.