Average Causal Effect in Observational Studies IV

Xinzhou Guo

HKUST

(Credited to Zhichao Jiang)

March 21, 2024

Ignorability $(Y(1), Y(0)) \perp Z \mid X$

Overlap $\eta \leq e(x) \leq 1-\eta \quad 0<\eta<1$

Implication of overlap assumption

Theorem (D'Amour et al., 2021)

If $\eta \leqslant e(\mathbf{X}_i) \leqslant 1-\eta$ for some $\eta \in (0, 1/2)$, then

$$\underbrace{\frac{1}{p}\sum_{k=1}^{p}\left|\mathbb{E}(X_{ik} \mid Z_i = 1) - \mathbb{E}(X_{ik} \mid Z_i = 0)\right|}$$

average covariate balance

$$\leqslant \frac{1}{p^{1/2}} C^{1/2}\left[e\lambda_{\max}^{1/2}\{\operatorname{cov}(\mathbf{X}_i \mid Z_i = 1)\} + (1-e)\lambda_{\max}^{1/2}\{\operatorname{cov}(\mathbf{X}_i \mid Z_i = 0)\}\right]$$

- $\lambda_{\max}^{1/2}\{\operatorname{cov}(\mathbf{X}_i \mid Z_i = z)\}$ is usually smaller than $O(p)$
- the right hand side converges to zero w.r.t. p if the components of X are not highly correlated $\rightsquigarrow$ X are nearly balanced in high dimensions

$\longrightarrow$ close to RCT

Average Covariate Balance $\lesssim$

$$\frac{1}{p^{\frac{1}{2}}} \left(\lambda_{max}^{\frac{r}{2}} \{ cov(X_i | Z_i = 1) \right.$$

$$\left. + \lambda_{max}^{\frac{r}{2}} \{ cov(X_i | Z_i = 0) \right]$$

$$cov(X_i | Z_i = 1) = cov(X_i | Z_i = 0)$$

$$= \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}_{p \times p}$$

$$\lambda_{max} = \sigma^2$$

Regression discontinuity design (RD design)

- An extreme case without overlap condition
  1. Sharp RD Design: treatment assignment is based on a deterministic rule $Z_i = \mathbf{1}(X_i \geqslant c)$ X_i: forcing variable
  2. Fuzzy RD Design
- Ignorability holds but overlap condition fails

$$e(X_i) = \begin{cases} 1 & \text{if } X \geqslant c \\ 0 & \text{if } X < c \end{cases}$$

- Originates from a study of the effect of scholarships on students' career plans (Thistlethwaite and Campbell. 1960. J. of Educ. Psychol)

Regression discontinuity design

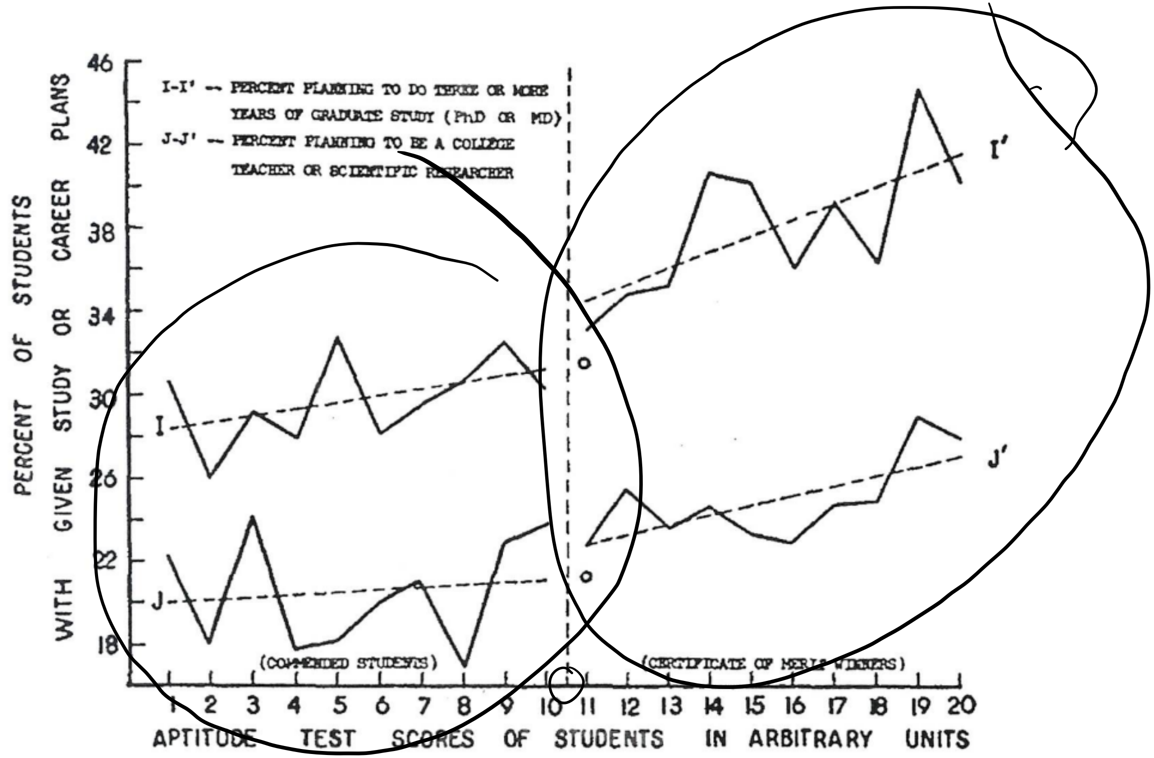


Fig. 3. Regression of study and career plans on exposure determiner.

Identification

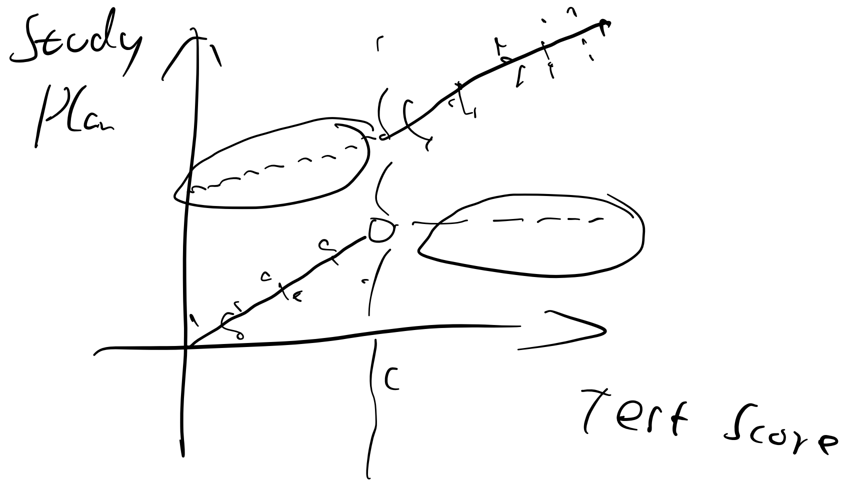
- Causal quantity of interest: $\mathbb{E}\{Y_i(1) - Y_i(0) \mid X_i = c\}$ – what if ACE?
- Assumption: $\mathbb{E}\{Y_i(z) \mid X_i = x\}$ is continuous in x for $z = 0, 1$
- Identification formula (why identifiable):

$$\mathbb{E}(Y_i(1) \mid X_i = c) = \lim_{x \downarrow c} \mathbb{E}(Y_i(1) \mid X_i = x) = \lim_{x \downarrow c} \mathbb{E}(Y_i \mid X_i = x)$$

$$\mathbb{E}(Y_i(0) \mid X_i = c) = \lim_{x \uparrow c} \mathbb{E}(Y_i(0) \mid X_i = x) = \lim_{x \uparrow c} \mathbb{E}(Y_i \mid X_i = x)$$

- Advantage: internal validity
- Disadvantage: external validity

$$\underbrace{ACE = E(Y(1) - Y(0))}$$



$$\underline{E(Y(1) \mid X = c)}$$

$$= \lim_{x \downarrow c} E(Y(1) \mid X = x)$$

$Z = I(X \geq c)$

$$= \lim_{x \downarrow c} E(Y(1) \mid X = x, Z = 1)$$

$$= \lim_{x \downarrow c} E(Y \mid X = x, Z = 1)$$

$$Y = Z\,Y(1) + (1 - Z)\,Y(0)$$

$$= \lim_{x \downarrow c} E(Y \mid X = x)$$

$E(Y(1) \mid X = c)$

by continuity

$= \lim_{x \downarrow c} E(Y(1) \mid X = x)$

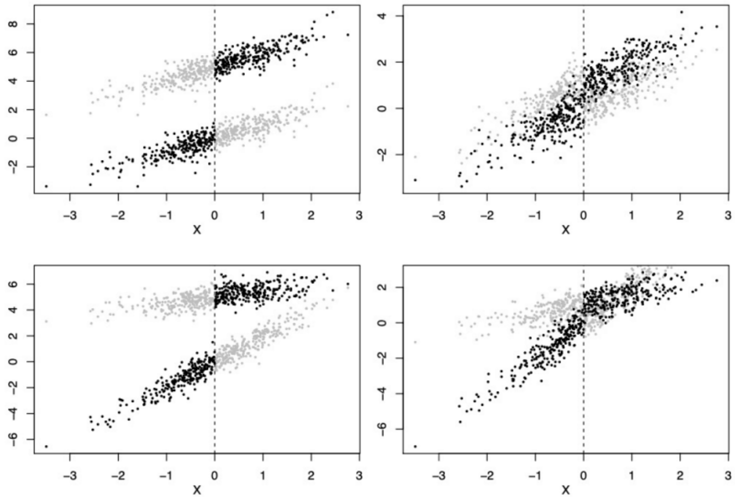
$e(x) = 1$

$= \lim_{x \downarrow c} E(Y \mid X = x)$

all the subjects whose $X > c$, receives treatment

Statistical inference: regression

- Graphical diagnostic sometimes shows the causal effect at the cutoff point



Statistical inference: regression

$$E_c\{Y(1) - Y(0) \mid X = c\}$$
$$= \lim_{x \downarrow c} E(Y \mid X = x) - \lim_{x \uparrow c} E(Y \mid X = x)$$

- Linear regression models

$$\mathbb{E}(Y_i \mid Z_i = 1, X_i) = \alpha_1 + \gamma_1 X_i$$
$$\mathbb{E}(Y_i \mid Z_i = 0, X_i) = \alpha_0 + \gamma_0 X_i$$

- $\widehat{\text{ACE}}(c) = (\hat{\alpha}_1 - \hat{\alpha}_0) + (\hat{\gamma}_1 - \hat{\gamma}_0)\, c$
- Equivalent implementation
  - coefficient of Z_i in regression of Y_i on $\{1, Z_i, X_i - c, Z_i (X_i - c)\}$
  - coefficient of Z_i in regression of Y_i on $\{1, Z_i, (X_i - c)^+, (X_i - c)^-\}$, where $(X_i - c)^+ = \max\{X_i - c, 0\}$ and $(X_i - c)^- = \min\{X_i - c, 0\}$

$$E(Y_i \mid Z_i = 1, X_i) - E(Y_i \mid Z_i = 0, X_i)$$

$$= E(Y_i \mid X_i = x, x \geq c)$$

$$- E(Y_i \mid X_i = x, x \leq c)$$

$$E(Y(1) - Y(0) \mid X = c)$$

$$= \lim_{x \downarrow c} E(Y \mid X, Z = 1)$$

$$- \lim_{x \uparrow c} E(Y \mid X, Z = 0)$$

$$= \alpha_1 + \gamma_1 c - (\alpha_0 + \gamma_0 c)$$

$$= \alpha_1 - \alpha_0 + (\gamma_1 - \gamma_0) c.$$

$$\lim_{x \downarrow c} E(Y \mid X = x) - \lim_{x \uparrow c} E(Y \mid X = c)$$

$E(Y \mid X = x)$ continuous w.r.t x

$$= \quad 0 \quad \neq 0$$

$E(Y(z) \mid X = x)$ is continuous

$$Y = \alpha + \beta X + \varepsilon$$

$$E(Y \mid X) = \alpha + \beta X$$

$E(Y_i \mid z_i = 1, X_i)$ $\quad$ $E(Y_i \mid z_i = 0, X_i)$

$Y \sim X$ for $z = 1$ $\quad$ $Y \sim X$ for $z = 0$

or for $X \geq c$ $\quad$ or for $X < c$

$$Y \sim z + (X_i - c) + z(X_i - c)$$

Statistical inference: local regression

- Limitation of using linear regression models
  - restricted functional form
  - equal contribution from all units
- Local linear regression (same h for both sides): better behavior at the boundary than other nonparametric regressions

$$\left(\hat{\alpha}_+, \hat{\beta}_+\right) = \underset{\alpha,\beta}{\operatorname{argmin}} \sum_{i=1}^{n} 1\left\{X_i > c\right\}\left\{Y_i - \alpha - (X_i - c)\,\beta\right\}^2 \cdot K\left(\frac{X_i - c}{h}\right)$$

$$\left(\hat{\alpha}_-, \hat{\beta}_-\right) = \underset{\alpha,\beta}{\operatorname{argmin}} \sum_{i=1}^{n} 1\left\{X_i < c\right\}\left\{Y_i - \alpha - (X_i - c)\,\beta\right\}^2 \cdot K\left(\frac{X_i - c}{h}\right)$$

- *rdrobust* package in R

$\hat{\alpha}_+ - \hat{\alpha}_-$

$$\int u\,k(u)\,du = 0 \qquad \int k(u)\,du = 1$$

$$k(u) = f_{N(0,1)}(u)$$

$$k(u) = \begin{cases} \frac{1}{2} & |u| \le 1 \\ 0 & |u| > 1 \end{cases}$$

$$k\left(\frac{x-c}{h}\right)$$

$$k\left(\frac{x-c}{h}\right) = 0 : x > c+h$$

$$< c-h$$

$$E\left(Y(1) - Y(0) \mid X = c\right)$$

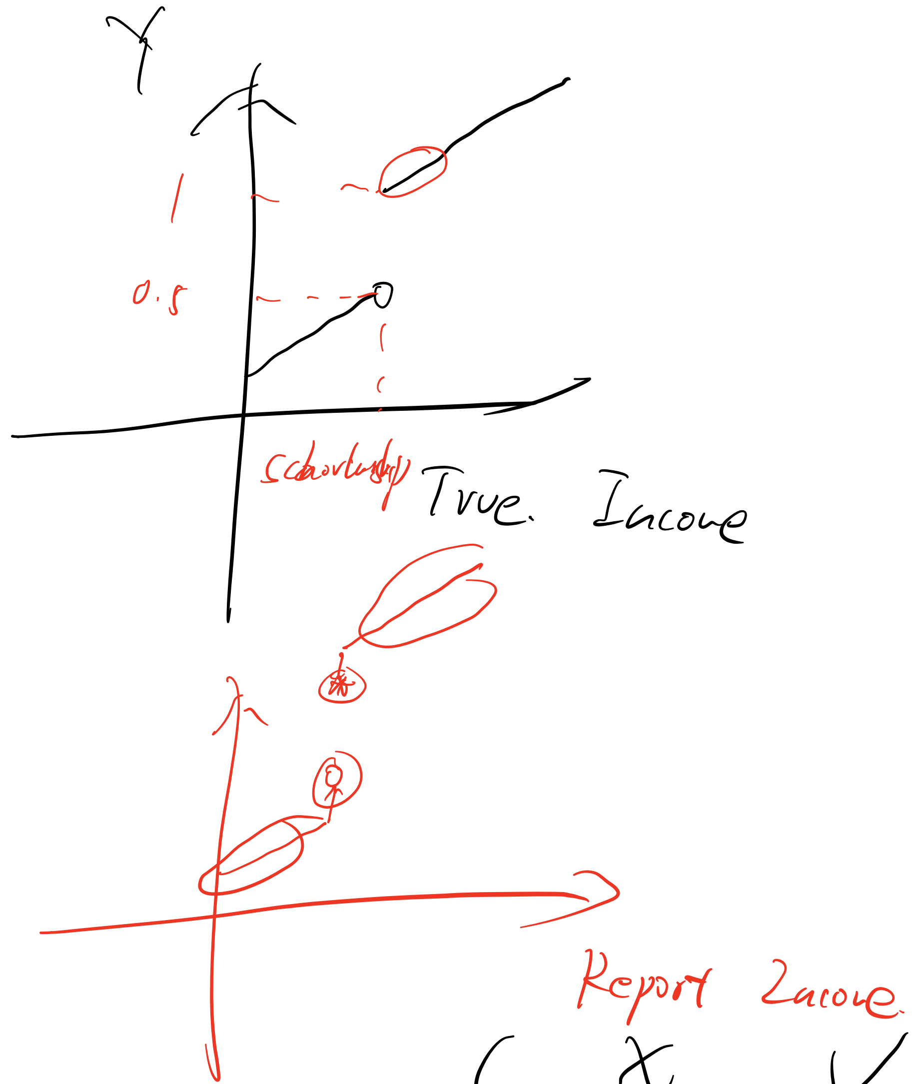
x

Checking the identification assumption of RD design

- RD design does NOT require the local randomization or "as-if random" assumption within a window:

$$\{Y_i(1), Y_i(0)\} \perp Z_i \mid c_0 \leqslant X_i \leqslant c_1$$

- Recall the key identification assumption: $\mathbb{E}\left(Y_i(z) \mid X_i = x\right)$ is continuous in x for $z = 0, 1$
- What does it mean to violate this assumption?
- If the treatment assignment mechanism is known to individuals, they may sort themselves into (or out of) the treatment group
- Examples of sorting or manipulation of forcing variable
  - reporting one's income just below the threshold to qualify for scholarships
  - students test score vs. teacher's assessment



$$E(W|X) = \begin{cases} X & \checkmark \\ c\,(c-\varepsilon \le X > c] + X & \times \end{cases}$$

$$E(Y(1)|X) \overset{W=X}{=} \beta + X\delta_1 \quad \checkmark$$

$$\overset{W\neq X}{=} E\left(E[Y(1)|W]\,|X\right)$$

$$= E(\beta + W\delta_1 | X) = \beta + \delta_1 E(W|X)$$

Continuity from nonrandom selection

- A simple model with X as the observed and W as the true.

$$Y(z) = \beta z + W\delta_1 + U$$

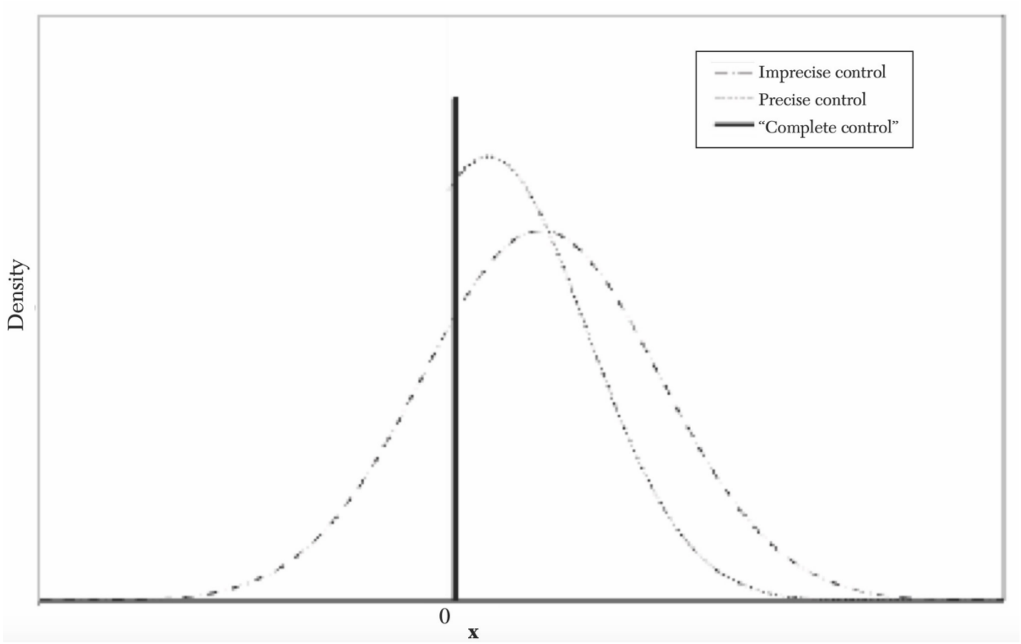
$$Z = \mathbf{1}(X \geqslant c), \quad X = W\delta_2 + V$$

- Bayes rule: $E(V|X) = X$ ✓

$$\Pr(W, U \mid X = x) = \frac{f(x \mid W, U)\Pr(W = w, U = u)}{f(x)}$$

- Individuals have imprecise control over X : $f(x \mid W, U)$ is continuous at the threshold $c \rightsquigarrow \mathbb{E}(Y(z) \mid X = x)$ is continuous at $X = c$
- See Lee and Lemieus (2010) "Regression Discontinuity Designs in Economics"

Continuity from nonrandom selection



Diagnostic tests

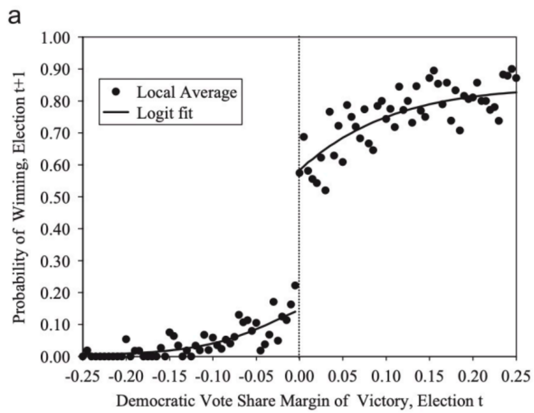
- Placebo test: check continuity for pre-treatment variables
  - expected not to have any effect
  - closely related to outcome of interest
  - lagged outcome
- Check the continuity of the density of the forcing variable
- No unexplainable discontinuity in the outcome-forcing variable relationship at values other than at the cutoff value

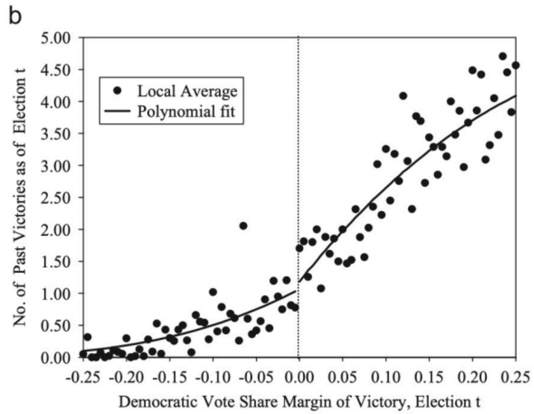
Incumbency advantage in the U.S. House (Lee, 2008)

- Re-election rate of the incumbent party is about 90% over the past 50 years
- Possible explanations
  - Representatives are using the privileges and resources of office to gain an "unfair" advantage over potential challengers.
  - Erikson (1971): incumbents, are, by definition, those politicians who were successful in the previous election. If what makes them successful is somewhat persistent over time, they should be expected to somewhat more successful when running for re-election.
- Question: is there any causal effect of being the incumbent party on the winning of the election
- Unit: congressional district
- Treatment Z_i : democrat is the incumbent party in a district
- Outcome Y_i : democrat winning the election

RDD analysis of the incumbency advantage

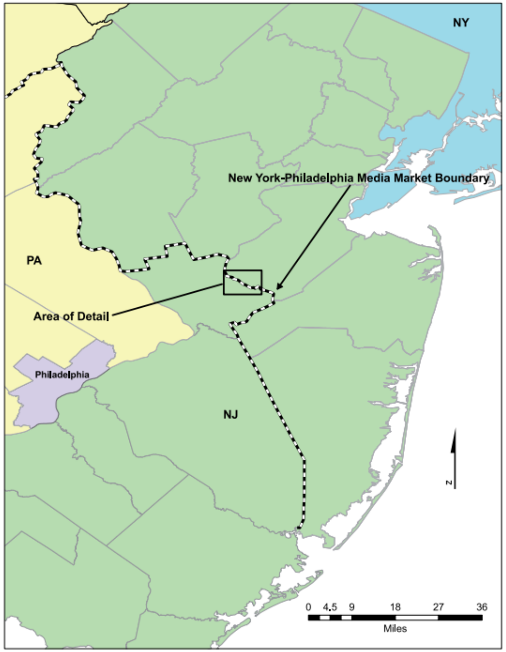
- Forcing variable X_i : vote share for Democrat - vote share for republican
- Deterministic rule: $Z_i = \mathbf{1}(X_i > 0)$
- Causal quantity: $\mathbb{E}\{Y_i(1) - Y_i(0) \mid X_i = 0\}$

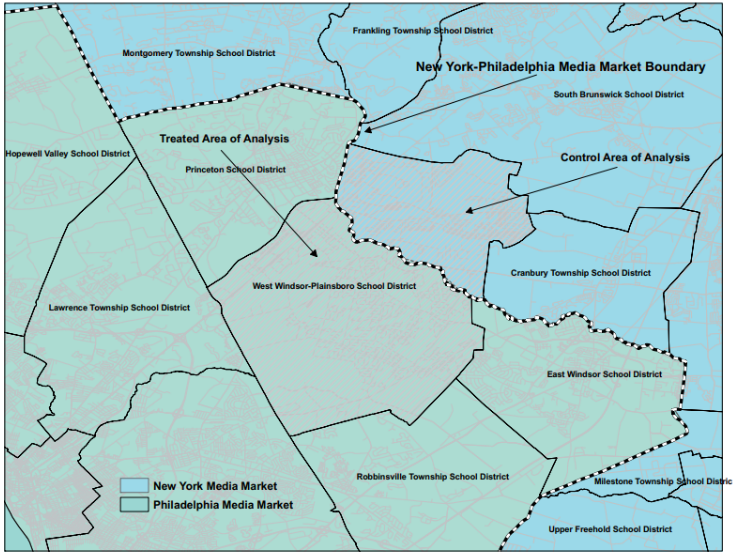




Geographical RD Design (Keele and Titiunik. 2015. Political Anal.)

- RD in two dimensions
- Example: media markets



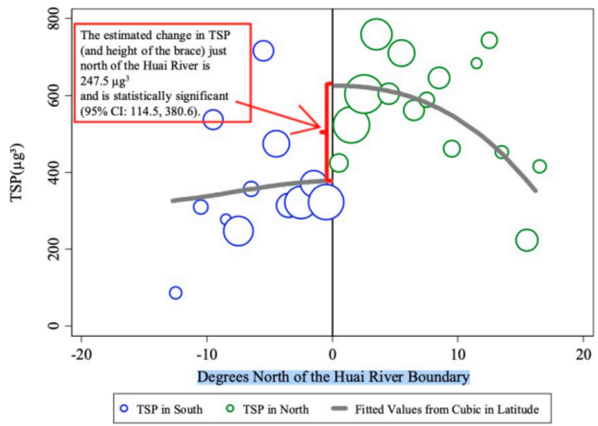


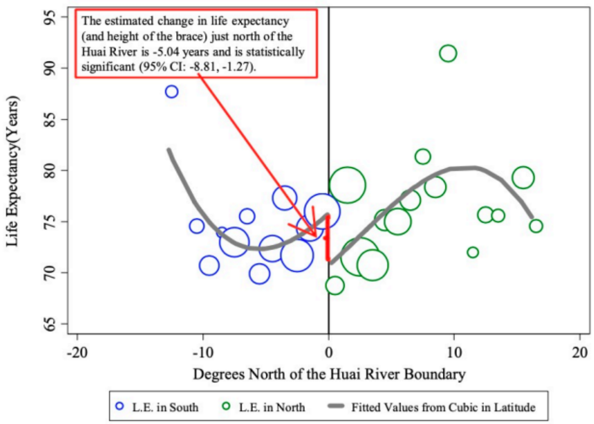
Effect of sustained exposure to air pollution on life expectancy (Chen et al. 2013. PNAS)

- Chinese government established free winter heating of homes and offices via the provision of free coal for fuel boilers as a basic right.
- Due to budgetary limitations, this right was only extended to areas located in North China, which is defined by the line formed by the Huai River and Qinling Mountain range

RD analysis

- Forcing variable X_i : Degrees north of the Huai river boundary
- Treatment: free heating provided
- Their conclusion: the policy that greatly increases TSPs air pollution is causing the 500 million residents of Northern China to lose more than 2.5 billion life years of life expectancy





Discussion

- Causal effect at the threshold $\neq$ Causal effect for the whole population
- Two outcomes: TSP and life expectancy
  - effect of having free heating on TSP at the threshold
  - effect of having free heating on life expectancy at the threshold
  - cannot draw conclusion about the effect of TSP on life expectancy
- Model sensitivity
- Other issues
  - migration of population
  - people living in different parts of China differ a lot

Summary

- Assessing overlap: plot propensity score for treatment and control group
- Regression discontinuity design
  - treatment is a deterministic function of covariates
  - Identify the ACE at the threshold
  - limited external validity $\rightsquigarrow$ extrapolation required for generalization

Suggested readings

- Regression discontinuity design
  - ANGRIST AND PISCHKE, Chapters 6.1
  - Cattaneo and Titiunik (2022) "Regression Discontinuity Designs"
- Statistical inference related
  - FAN AND GiJBELS "Local Polynomial Modelling and Its Applications"
  - bandwidth selection : Imbens and Kalyanaraman (2012) "Optimal bandwidth choice for the regression discontinuity estimator"
- Continuity versus local randomization
  - de la Cuesta and Imai (2016) "Misunderstandings about the Regression Discontinuity Design in Close Elections Dataverse"

Suggested readings

- Interrupted time series
  - Miritrix (2022) "Using Simulation to Analyze Interrupted Time Series Designs"
- Non-extreme case: overlap violated for some values of X
  - outcome modeling for extrapolation
  - overlap weights: Li and Zaslavsky (2018) "Balancing covariates via propensity score weighting"
  - stochastic intervention: Kennedy (2018) "Nonparametric Causal Effects Based on Incremental Propensity Score Interventions"