Regression Discontinuity Designs

Kosuke Imai

Harvard University

STAT186/GOV2002 CAUSAL INFERENCE

Fall 2018

Observational Studies

- In many cases, we cannot randomize the treatment assignment
  - ethical constraints
  - logistical constraints
- But, some important questions demand empirical evidence even though we cannot conduct randomized experiments!
- Designing observational studies $\rightsquigarrow$ find a setting where credible causal inference is possible
- Key $=$ Knowledge of treatment assignment mechanism
- Regression discontinuiety design (RD Design):
  1. Sharp RD Design: treatment assignment is based on a *deterministic* rule
  2. Fuzzy RD Design: encouragement to receive treatment is based on a deterministic rule
- Originates from a study of the effect of scholarships on students' career plans (Thistlethwaite and Campbell. 1960. *J. of Educ. Psychol*)

Regression Discontinuity Design

- Idea: Find an arbitrary cutpoint c which determines the treatment assignment such that $T_i = \mathbf{1}\{X_i \geq c\}$
- Close elections as RD design (Lee et al. 2004. *Q. J. Econ*):

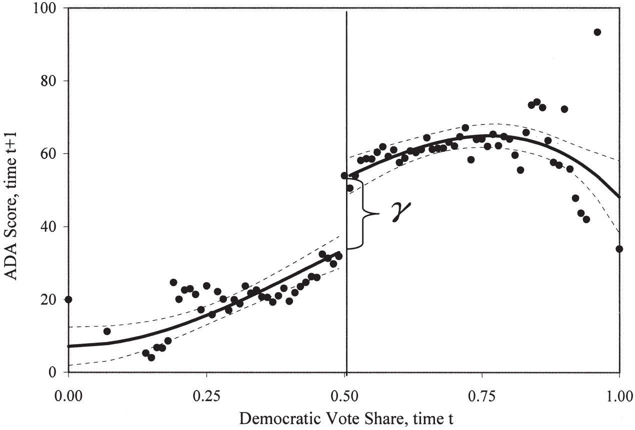


FIGURE I
Total Effect of Initial Win on Future ADA Scores: γ

Identification

- Estimand:
$$\mathbb{E}(Y_i(1) - Y_i(0) \mid X_i = c)$$
- Assumption: $\mathbb{E}(Y_i(t) \mid X_i = x)$ is continuous in x for $t = 0, 1$
  - deterministic rather than stochastic treatment assignment
  - violation of the overlap assumption: $0 < \Pr(T_i \mid X_i = x) < 1$ for all x
  - RD design is all about extrapolation
- Regression modeling:
$$\begin{aligned}
\mathbb{E}(Y_i(1) \mid X_i = c) &= \lim_{x \downarrow c} \mathbb{E}(Y_i(1) \mid X_i = x) = \lim_{x \downarrow c} \mathbb{E}(Y_i \mid X_i = x) \\
\mathbb{E}(Y_i(0) \mid X_i = c) &= \lim_{x \uparrow c} \mathbb{E}(Y_i(0) \mid X_i = x) = \lim_{x \uparrow c} \mathbb{E}(Y_i \mid X_i = x)
\end{aligned}$$
- Advantage: internal validity
- Disadvantage: external validity
- Make sure nothing else is going on at $X_i = c$

Analysis Methods under the RD Design

- Simple linear regression within a window
  - How should we choose a window in a principled manner?
  - How should we relax the functional form assumption?
  - higher-order polynomial regression $\rightsquigarrow$ not robust to outliers
- Local linear regression (same h for both sides): better behavior at the boundary than other nonparametric regressions

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

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

- Weighted regression with a kernel function of one's choice:
  - uniform kernel: $K(u) = \frac{1}{2}\mathbf{1}\{|u| < 1\}$
  - triangular kernel: $K(u) = (1 - |u|)\mathbf{1}\{|u| < 1\}$

Optimal Bandwidth (Imbens and Kalyanaraman. 2012. *Rev. Econ. Stud.*)

- Choose the bandwidth by minimizing the MSE:

$$\begin{aligned}
\text{MSE} &= \mathbb{E}[\{(\hat{\alpha}_+ - \hat{\alpha}_-) - (\alpha_+ - \alpha_-)\}^2 \mid \mathbf{X}] \\
&= \mathbb{E}\{(\hat{\alpha}_+ - \alpha_+)^2 \mid \mathbf{X}\} + \mathbb{E}\{(\hat{\alpha}_- - \alpha_-)^2 \mid \mathbf{X}\} \\
&\quad -2 \cdot \mathbb{E}(\hat{\alpha}_+ - \alpha_+ \mid \mathbf{X}) \cdot \mathbb{E}(\hat{\alpha}_- - \alpha_- \mid \mathbf{X}) \\
&= (\text{Bias}_+ - \text{Bias}_-)^2 + \text{Variance}_+ + \text{Variance}_-
\end{aligned}$$

- Bias and variance of local linear regression estimator at the boundary:

$$\text{Bias} = \mathbb{E}(\hat{m}(0) \mid \mathbf{X}) - m(0), \quad \text{Variance} = \mathbb{V}(\hat{m}(0) \mid \mathbf{X})$$

where $m(x) = \mathbb{E}(Y_i \mid X_i = x)$, $\hat{m}(x) = \hat{\alpha}(x)$, and

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

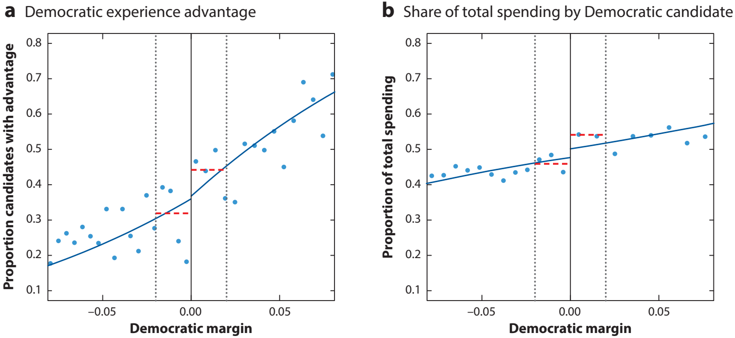
- Refinements, e.g., bias correction (Calonico et al. 2014. *Econometrica*)

The “As-if Random” Assumption

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

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i \mid c_0 \leq X_i \leq c_1$$

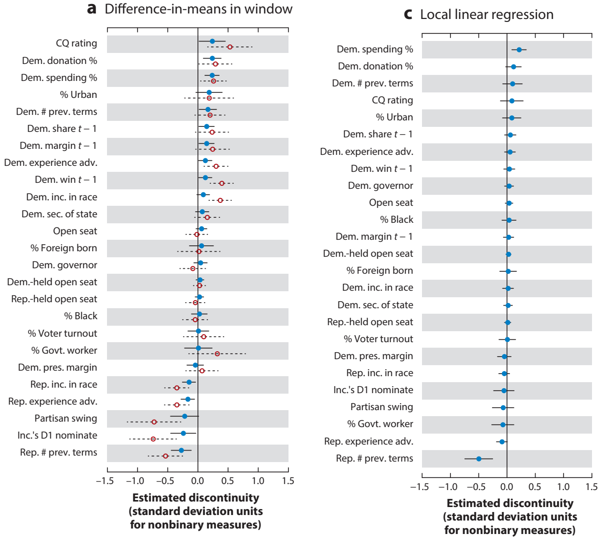
- The “as-if random” assumption implies zero slope of regression lines $\rightsquigarrow$ difference-in-means within the window
- The assumption may be violated regardless of the window size



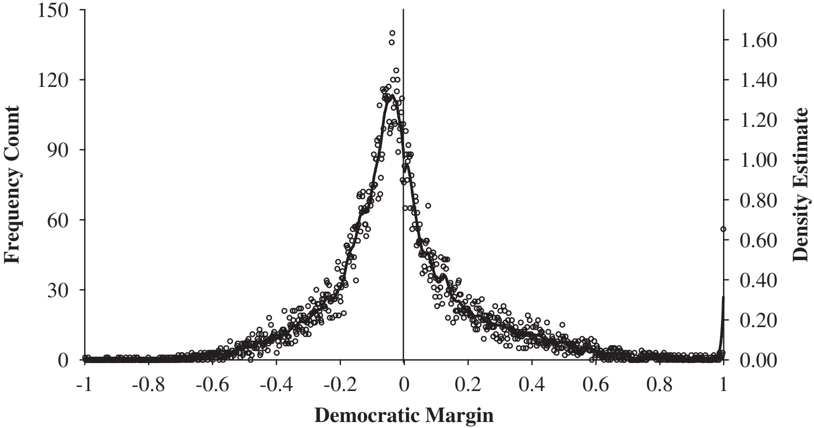
Close Elections Controversy (de la Cuesta and Imai. 2016. *Annu. Rev. Political Sci*)

Sorting?

1. Pre-election behavior or characteristics of candidates, e.g., resource advantages ⇝ steep slope
2. Post-election advantages in vote tallying, e.g., election fraud ⇝ sorting



Density Test of Sorting (McCrary. 2008. *J. Econom.*)



1. Create histogram with a selected bin size
2. Fit local linear regression to bin midpoints to smooth the histogram
3. Estimate the difference in the logged histogram height at the threshold

Placebo Test

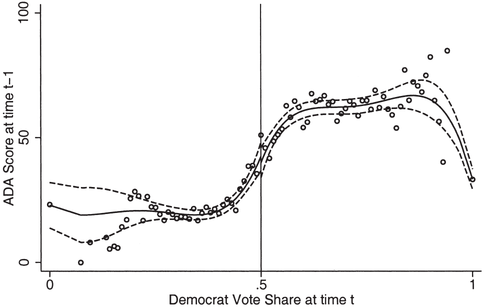


FIGURE V

Specification Test: Similarity of Historical Voting Patterns between Bare Democrat and Republican Districts

- What is a good placebo?
  1. expected not to have any effect
  2. closely related to outcome of interest
- Lagged outcome $\rightsquigarrow$ future cannot affect past
- Interpretation: failure to reject the null $\neq$ the null is correct

Fuzzy RD Design (Hahn et al. 2001. *Econometrica*)

- Sharp regression discontinuity design: $T_i = \mathbf{1}\{X_i \geq c\}$
- What happens if we have noncompliance?
- Forcing variable as an instrument: $Z_i = \mathbf{1}\{X_i \geq c\}$
- Potential outcomes: $T_i(z)$ and $Y_i(z, t)$
- Assumptions
  1. Monotonicity: $T_i(1) \geq T_i(0)$
  2. Exclusion restriction: $Y_i(0, t) = Y_i(1, t)$
  3. $\mathbb{E}(T_i(z) \mid X_i = x)$ and $\mathbb{E}(Y_i(z, T_i(z)) \mid X_i = x)$ are continuous in x
- Estimand:

$$\mathbb{E}(Y_i(1, T_i(1)) - Y_i(0, T_i(0)) \mid Complier, X_i = c)$$

- Estimator:

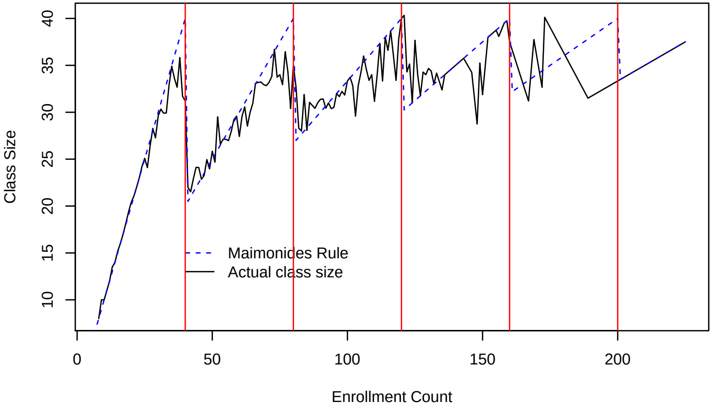
$$\frac{\lim_{x \downarrow c} \mathbb{E}(Y_i \mid X_i = x) - \lim_{x \uparrow c} \mathbb{E}(Y_i \mid X_i = x)}{\lim_{x \downarrow c} \mathbb{E}(T_i \mid X_i = x) - \lim_{x \uparrow c} \mathbb{E}(T_i \mid X_i = x)}$$

- Disadvantage: external validity

Class Size Effect (Angrist and Lavy. 1999. *Q. J. Econ*)

- Effect of class-size on student test scores
- Maimonides' Rule: Maximum class size $= 40$

$$f(z) \;=\; \frac{z}{\left\lfloor \frac{z-1}{40} \right\rfloor + 1}$$



Empirical Analysis

- Y_i: class average verbal test score
- Window size: w
- Construction of forcing variable:

$$X_i = \begin{cases} 40 - Z_i & \text{if } 40 - w/2 \leq Z_i \leq 40 + w/2 \\ 80 - Z_i & \text{if } 80 - w/2 \leq Z_i \leq 80 + w/2 \\ \vdots & \qquad\qquad \vdots \end{cases}$$

- Linear models (cluster standard errors by schools):

$$\begin{aligned} T_i &= \alpha_1 \times I\{X_i >= 0\} + \beta_1 X_i + \gamma_1 X_i \times I\{X_i >= 0\} + \epsilon_{1i} \\ Y_i &= \alpha_2 \times I\{X_i >= 0\} + \beta_2 X_i + \gamma_2 X_i \times I\{X_i >= 0\} + \epsilon_{2i} \end{aligned}$$

  where $\hat{\alpha}_1 = -7.90$ (s.e. $= 1.90$) and $\hat{\alpha}_2 = -0.056$ (s.e. $= 2.08$)
- Two-stage least squares estimate: est. $= 0.007$ (s.e. $= 0.261$)

Summary

- Observational studies $\rightsquigarrow$ treatment assignment is not random
- "Design" observational studies for credible causal inference

- Sharp regression discontinuity designs:
  - deterministic (rather than stochastic) treatment assignment rule
  - continuity assumption $\rightsquigarrow$ no sorting
  - does not require "as-if random" assumption
  - limited external validity $\rightsquigarrow$ extrapolation required for generalization
  - incumbency effects controversy (Eggers et al. 2015. *Am. J. Political Sci*)
  - Importance of placebo tests
- Fuzzy regression discontinuity designs: noncompliance

- Suggested readings:
  - ANGRIST AND PISCHKE, Chapter 6
  - IMBENS AND LEMIEUX. (2008). *J. of Econom*