Session 12: Regression discontinuity

MGT 581 | Introduction to econometrics

Michaël Aklin

PASU Lab | EPFL

Last time...

- Difference-in-differences
- Panel data
- Two-way fixed effects models

Today:

- Regression discontinuity design

Readings:

Regression discontinuity design

- Regression discontinuity: another type of quasi-experiment
- Originally from the 1960s but picked up in late 1990s (Thistlethwaite and Campbell 1960; Lee and Lemieux 2010)
- General idea of quasi-experiments: can we find settings in which receiving treatment or control is **as if random**?
- Here: assume you could find cases where being in Treatment or Control is almost a coin toss
- **Discontinuity**: "frontier" or "cutoff" at which units either end up T or C (**sharp RDD**) or at which $Pr(D = 1)$ changes (**fuzzy RDD**)
- Assumption: very close to the discontinuity, whether you end up T or C is near-random and thus like an experiment

Illustration

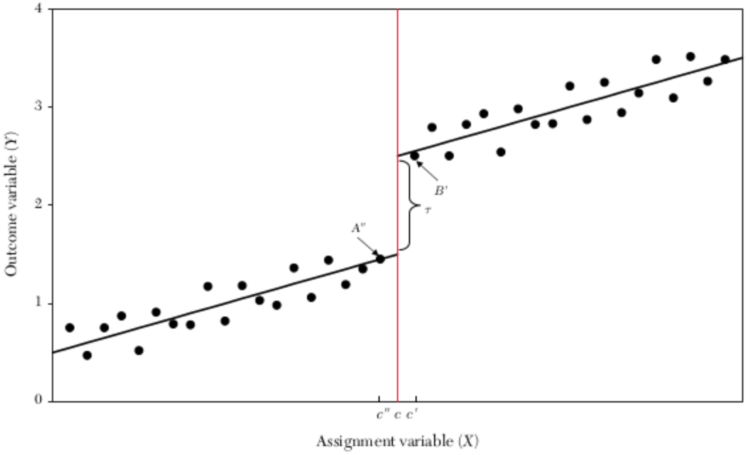

Figure 1. Simple Linear RD Setup

Figure 1: Here: X is the assignment/forcing variable (below: Z). c is the discontinuity (below: z_0). Source: OSE data science.

- Idea: at the discontinuity, units are identical (in expectation) on either side *except* for the treatment D they received
- Assignment to treatment is a function of an **assignment** or **forcing** variable
- Three primary identifying assumptions (Hahn, Todd, and Van der Klaauw 2001).
- Notation: treatment D, forcing variable Z, and defining the discontinuty to occur at cutoff value z_0

- Define:

$$D^{+} \equiv \lim_{z \to z_0^{+}} E[D_i = 1 | z_i = z],$$
$$D^{-} \equiv \lim_{z \to z_0^{-}} E[D_i = 1 | z_i = z],$$

- Assumptions:

1 : The limits above exist
$2 : D^{+} \neq D^{-}$
$3 : E[Y | D = 0, z_i = z]$ is continuous at z_0

- Then, RD estimator:

$$\beta = \frac{Y^{+} - Y^{-}}{D^{+} - D^{-}}$$

- Note: Y^+ and Y^- are defined as:

$$Y^+ \equiv \lim_{z \to z_0^+} E[Y_i | z_i = z],$$
$$Y^- \equiv \lim_{z \to z_0^-} E[Y_i | z_i = z],$$

- The RD is the slope of the treatment effect at the discontinuity (change in Y divided by change in D)
- The RD estimator is defined (numerator is not 0 by assumption 2)

- Rationale: the set of units just “below” the discontinuity is a good **comparison group** for the set of units just “above”. We say that $\mathbf{X}$ is **continuous** or **balanced** at z_0
- Thus: no confounder/backdoor channel
- Advantage: **no need to adjust** for covariates $\mathbf{X}$, since they should not vary (much) on average at the discontinuity
- Advantage: no worry about functional forms (same reason: no variation on control variables $\mathbf{X}$)

Some plausible RDDs:

- Effect of being elected to power: those just below and those just above 50% (Eggers and Hainmueller 2009)
- Effect of classroom size: those just below and just above number needed to split class in two (Angrist and Lavy 1999)
- Majority age (comparing people at age 17.9 vs. 18.1: they could plausibly be the same *except* that one group has the right to vote)

Example

FIGURE 4. Regression Discontinuity Design: Effect of Serving in House of Commons on Wealth at Death

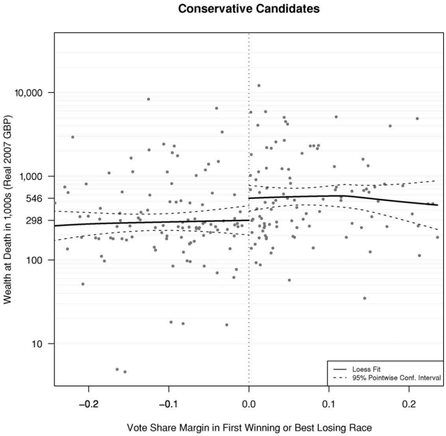

Figure 2: Effect of being elected Conservative Member of Parliament (MP) in the UK on their long-term wealth. Source: Eggers and Hainmueller (2009)

Example

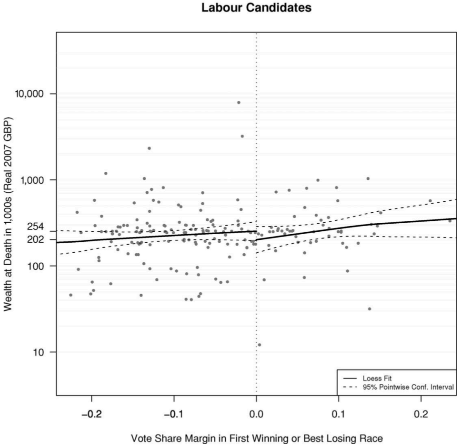

Figure 3: Effect of being elected Labour Member of Parliament (MP) in the UK on their long-term wealth. Source: Eggers and Hainmueller (2009)

Identifying assumption

- Underlying assumption: no selection on either side of the discontinuity (Hahn, Todd, and Van der Klaauw 2001)
- Check otherwise likely control variables: should not differ at the threshold (McCrary 2008)
    - For instance: effect of *being elected* on *wealth*: on average, treated and control should have the same age, educational levels, etc.

In practice:

- Identify an exogenous discontinuity
- Start with a plausible bandwidth/sphere around the discontinuity
    - Eg close elections are those within $+/-1\%$ of winning
    - Alternatively: let an algorithm decide (often preferable)
- Collect data within these boundaries
- Estimate with OLS

$$Y_i = \alpha + \delta \text{Forcing}_i + \gamma D_i + \beta D_i \cdot \text{Forcing}_i$$

- β is the estimate of treatment effect (the 'jump' in Y at z_0)
- Note: interaction term is needed to let slope differ at discontinuity

Challenge

- How do you model the *forcing* variable?
- Could in principle add higher polynomials: 2nd, 3rd, etc.

$$Y_i = \alpha + \delta_1 \mathsf{F}_i + \gamma D_i + \beta_1 D_i \cdot \mathsf{F}_i + \delta_2 \mathsf{F}_i^2 + \beta_2 D_i \cdot \mathsf{F}_i^2$$

- Gelman and Imbens (2019): keep it simple and don't go above quadratic

Conclusion

- RDD is a powerful solution to endogeneity and issues such as functional form
- Idea is simple: at discontinuity, units are otherwise identical
- Limited number of cases where discontinuities exist

References

Angrist, Joshua, and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *The Quarterly Journal of Economics* 114 (2).

Eggers, Andrew C, and Jens Hainmueller. 2009. "MPs for Sale? Returns to Office in Postwar British Politics." *American Political Science Review* 103 (4): 513–33.

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): 447–56.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (1): 201–9.

Lee, David S, and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2): 698–714.

Thistlethwaite, Donald L, and Donald T Campbell. 1960. "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment." *Journal of Educational Psychology* 51 (6): 309.