Regression Discontinuity Design II

Ye Wang
University of North Carolina at Chapel Hill

Linear Methods in Causal Inference
POLI784

Review

- In sharp RDD, the treatment is determined by the value of the running variable.
- We can only identify the causal effect for units at the cutoff.
- We assume the continuity of the expected outcome across the cutoff and rely on kernel regression for estimation.
- It is necessary to evaluate the assumption of continuity with placebo outcomes, placebo cutoffs, and the McCrary test.
- Bandwidth is selected to minimize the MSE of the intercept estimates.
- Bias correction ensures that the estimate converges to a normal distribution.

Fuzzy RDD

- Today we are going to discuss several variants of the sharp RDD.
- The first variant is the fuzzy RDD, in which D_i is affected by Z_i in the following way:

$$D_i = \begin{cases} D_i(1) \text{ if } Z_i \geq 0 \\ D_i(0) \text{ if } Z_i < 0 \end{cases}$$

- $D_i(1)$ may not be 1 and $D_i(0)$ may not be 0.
- In other words, we now have non-compliance in the ideal experiment.
- $\mathbf{1}\{Z_i \geq 0\}$ is an instrument of D_i.

Fuzzy RDD

- $\mathbf{1}\{Z_i \geq 0\}$ should satisfy all the requirements for an instrument.
- Then, we can identify the treatment effect on the compliers when $Z_i = 0$:

$$\tau_{FRD} = \frac{E[Y_i(1) - Y_i(0)|Z_i = 0]}{E[D_i(1) - D_i(0)|Z_i = 0]}.$$

- Naturally, we can estimate the quantity with

$$\hat{\tau}_{FRD} = \frac{\hat{\mu}_{Y+} - \hat{\mu}_{Y-}}{\hat{\mu}_{D+} - \hat{\mu}_{D-}}.$$

where all the four intercepts are estimated via local regression as in the previous lecture.

Fuzzy RDD

- $\hat{\tau}_{FRD}$ is a Wald estimator and approximately a linear combination of two sharp RDD estimates (one for Y and the other for D).
- Bias correction is also necessary for the estimate to be asymptotically normal.
- It must be conducted for both sharp RDD estimators.
- Calonico, Cattaneo, and Titiunik (2014) show that

$$\frac{\hat{\tau}_{FRD}^{bc} - \tau_{FRD}}{\sqrt{Var[\hat{\tau}_{FRD}^{bc}]}} \to N(0,1)$$

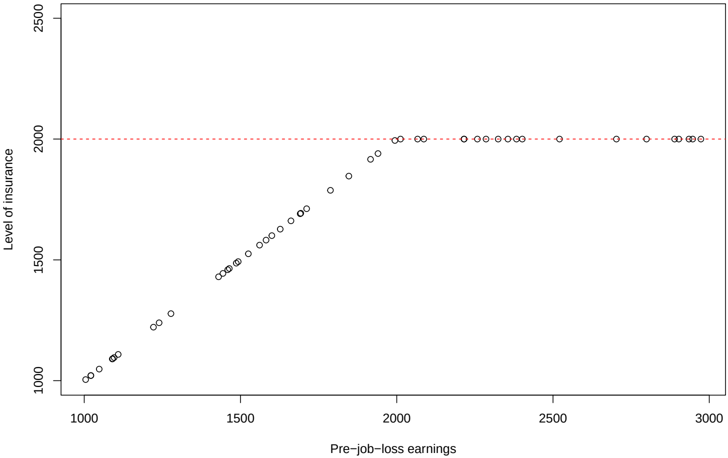
under the same condition: $N \min\{b^5, h^5\} \max\{b^2, h^2\} \to 0$.

Kink

- In a kink design, the treatment's level may not vary across the cutoff, but its rate of change might.
- For example, Y_i is the duration of unemployment, D_i is the level of unemployment insurance, and Z_i is one's pre-job-loss earnings.
- We want to estimate how Y_i responses to the change of D_i.
- Clearly, the correlation between Y_i and D_i is inevitably driven by confounders.

Kink

- Suppose the level of unemployment insurance equals to one's pre-job-loss earnings but there is a cap: $D_i = \min\{Z_i, 2000\}$:



Kink

- We can exploit this kink in the treatment to identify the response of the outcome to the treatment.
- Notice that the treatment is a deterministic and continuous function of running variable, $D = d(z)$.
- Suppose there is also a kink in the outcome across the cutoff, then the quantity of interest is

$$\tau_{SKRD} = \frac{\mu_+^{(1)} - \mu_-^{(1)}}{d_+^{(1)} - d_-^{(1)}}$$

where $d_+^{(1)} = \lim_{z \to 0_+} \frac{d(d(z))}{dz}$ ($D_-^{(1)}$ is similarly defined).

Kink

- The denominator is not at random and can be directly obtained.
- The numerator is the difference between two slopes and they can be similarly estimated via local regression.
- The asymptotic distribution of the numerator can be derived based on the theory in Calonico, Cattaneo, and Titiunik (2014):

$$\frac{\hat{\tau}^{bc}_{SKRD} - \tau_{SKRD}}{\sqrt{Var[\hat{\tau}^{bc}_{SKRD}]}} \rightarrow N(0, 1)$$

  as long as $N \min\{b^7_N, h^7_N\} \max\{b^2_N, h^2_N\} \rightarrow 0$.
- In theory, we can estimate higher order derivatives but their substantive meanings are unclear.
- We also have the fuzzy kink design where there is non-compliance in the treatment.

Multidimensional RDD

- So far, we have assumed that the running variable Z_i is uni-dimensional.
- But in practice, Z_i could be the location of unit i on a geography, which is decided by both its latitude and longitude.
- L. J. Keele and Titiunik (2015) show that the identification assumption is that $\mu(\cdot)$ is continuous along both dimensions across the cutoff (border).
- First, we select a series of locations along the border.
- Next, we apply the local regression estimator to generate one estimate for each location.
- Each unit is weighted by its distance to this location.
- Finally, we aggregate the estimates by re-weighting them with the density at the corresponding location.
- Another option is to use the minimum distance to the border as a uni-dimensional running variable.
- This approach 1. loses information and 2. prevents us from seeing the heterogeneity in treatment effects.

Multidimensional RDD

- It is usually harder to clarify what the treatment captures when Z_i is multi-dimensional (Dell, Lane, and Querubin 2018).
- Many things change together across the border of an administrative unit.
- Sorting (immigration) and autocorrelation are more likely to occur in the spatial setting.
- L. Keele, Titiunik, and Zubizarreta (2015) propose that researchers first match units on both covariates and their geographic locations and then generate RDD estimates on the matched sample.

RDD with a discrete running variable

- The classic theory developed in Calonico, Cattaneo, and Titiunik (2014) runs into difficulties when the running variable is discrete (e.g., age or time).
- By definition, there cannot be more observations around the cutoff point as N increases.
- Remember that

$$\frac{\hat{\tau}_{SRD} - \tau_{SRD}}{\sqrt{Var[\hat{\tau}_{SRD}]}} = \frac{\hat{\tau}_{SRD} - E[\hat{\tau}_{SRD}]}{\sqrt{Var[\hat{\tau}_{SRD}]}} + \frac{E[\hat{\tau}_{SRD}] - \tau_{SRD}}{\sqrt{Var[\hat{\tau}_{SRD}]}}.$$

- Now, the second term does not converge to zero even with bias correction.

RDD with a discrete running variable

- Kolesár and Rothe (2018) provide a finite-sample confidence interval in this case.
- We use the same kernel regression and variance estimators.
- Instead of correcting the bias, we adjust the estimate with an estimate of the maximal bias.
- The CI will be "honest" as it adapts to the worst possible scenario.
- It can be implemented with the package *RDHonest*.
- This approach usually leads to very conservative inferences.
- Ghosh, Imbens, and Wager (2025) argue that we can infer the maximal bias using the bias-correction estimator if the treatment effect is linear in the running variable.

References I

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Dell, Melissa, Nathan Lane, and Pablo Querubin. 2018. "The Historical State, Local Collective Action, and Economic Development in Vietnam." *Econometrica* 86 (6): 2083–2121.

Ghosh, Aditya, Guido Imbens, and Stefan Wager. 2025. "PLRD: Partially Linear Regression Discontinuity Inference." *arXiv Preprint arXiv:2503.09907*.

Keele, Luke J, and Rocio Titiunik. 2015. "Geographic Boundaries as Regression Discontinuities." *Political Analysis* 23 (1): 127–55.

Keele, Luke, Rocio Titiunik, and José R Zubizarreta. 2015. "Enhancing a Geographic Regression Discontinuity Design Through Matching to Estimate the Effect of Ballot Initiatives on Voter Turnout." *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 223–39.

References II

Kolesár, Michal, and Christoph Rothe. 2018. "Inference in Regression Discontinuity Designs with a Discrete Running Variable." *American Economic Review* 108 (8): 2277–2304.