Geographic Difference-in-Discontinuities

Kyle Butts*

NOVEMBER 2, 2021

A recent econometric literature has critiqued the use of regression discontinuities where administrative borders serve as the 'cutoff'. Identification in this context is difficult since multiple treatments can change at the cutoff and individuals can easily sort on either side of the border. This note extends the difference-in-discontinuities framework discussed in Grembi et al. (2016) to a geographic setting. The paper formalizes the identifying assumptions in this context which will allow for the removal of time-invariant sorting and multiple treatments similar to the difference-in-differences methodology.

Keywords: difference-in-discontinuities, spatial econometrics, sorting, compound treatment, causal inference

JEL Classification Number: C01, R15, R58

*University of Colorado Boulder, Economics Department (kyle.butts@colorado.edu)

1 — Introduction

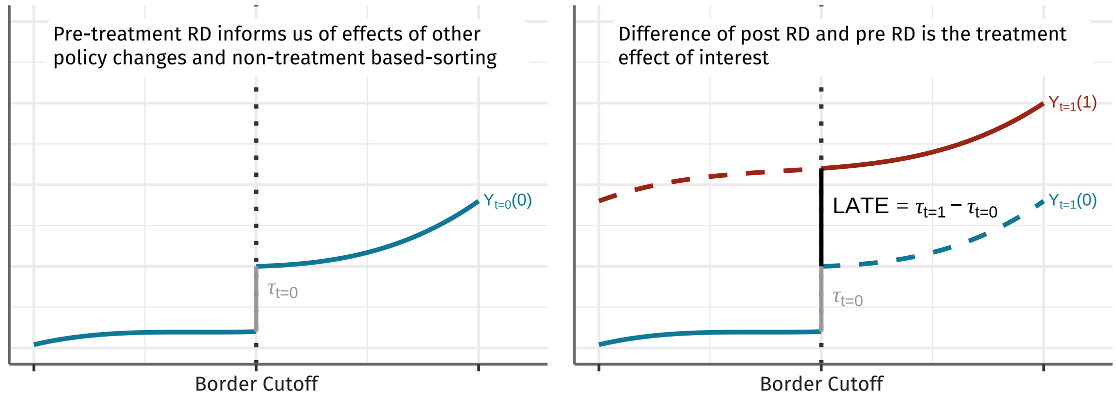
An increasingly popular estimation strategy involves using administrative borders as cutoffs in a regression discontinuity (RD) setting where the 'running variable' is the distance to the border. The purpose of using observations close to the border 'cutoff' is to try and better match treated and control units based on unobservable characteristics. Identification using the standard RD continuity assumption is problematic because many laws and institutions change discontinuously at a border (i.e. compound treatment) and people choose to sort on either side of a border (i.e. sorting around cutoff), leading to important differences between units in close geographic proximity even in the counterfactual world without treatment.[1]

The intuition of the difference-in-discontinuities design is very similar to the difference-in-differences design. A pre-treatment RD identifies time-invariant effects of other laws as well as the discontinuity in outcomes due to time-invariant sorting. A post-treatment RD identifies those previous two discontinuities plus the one caused by the treatment of interest. The difference between the two identifies the treatment effect. In this note, I extend the difference-in-discontinuities identification strategy formalized in Grembi et al. (2016); Eggers et al. (2018) to the context of geographic RDs and discuss the particular identifying assumptions needed for the above identification sketch to be true when using a geographic RD.

Figure 1 shows a stylized example of this methodology. The left panel shows a discontinuity at the border cutoff that exists before treatment. This could be due to other policies changing at the border or sorting due to other reasons. The right panel shows the treated and untreated potential outcomes in the post-period. The key idea behind the difference-in-discontinuities estimator is that the pre-period discontinuity can be estimated and removed from the second-period discontinuity to estimate the treatment

1. Identification through randomization local to the cutoff does not make sense in the geographic context because that would require people to randomly be located on either side of the border.

Figure 1 — Example of Difference-in-Discontinuities Identification



Notes: This figure shows a stylized example of identification in the difference-in-discontinuities setting.

effect so long as the magnitude of the discontinuity remains constant between periods.

I contribute to the econometric literature on RD in three ways. First, I contribute to the nascent literature formalizing the difference-in-discontinuities identification strategy. The results of the previous papers only consider the case of compound treatment where multiple treatments occur at a cutoff. Grembi et al. (2016); Eggers et al. (2018) analyze the situation where treatment is assigned based on population thresholds. Galindo-Silva et al. (2021); Millán-Quijano (2020) study a fuzzy-discontinuity setting where treatment probability changes discontinuously at an age cutoff but only among 'compliers'. This paper formalizes the difference-in-discontinuities in geographic settings.

The second contribution is that I extend on the work of Keele and Titiunik (2015) who formalize identification with geographic RDs into the *geographic difference-in-discontinuities* setting. The authors raise the problem of sorting on either side of the boundary as well as multiple laws changing discontinuously at the boundary and propose stringent assumptions to avoid these problems in the cross-section. This paper uses the difference-in-discontinuities methodology which provides a solution to these problems under arguably less stringent assumptions by leveraging the panel nature of data to estimate time-invariant sorting and effects of other policy/institution changes.

Last, I show that estimation of difference-in-discontinuities with panel data can be done by running RD on outcomes that have been first-differenced. This allows for the use of modern advancements in the estimation and inference of regression-discontinuities.[2] In cases where panel data is not available, then the local regression framework proposed in Grembi et al. (2016) can be used.

2 — Methods

2.1. Traditional RD Identification

Before introducing the difference-in-discontinuities method, I first review geographic RD to highlight difficulties in cross-sectional identification. I consider the standard context a random independent and identically distributed sample of units $i \in \{1, \dots n\}$. There is a running variable D_i that measures distance to the border of a treated area. Without loss of generality, the distance is normalized to zero with positive distances being within the treatment area.[3] The observed outcome is modeled by

$$y_i = f(D_i) + \tau(D_i)\mathbf{1}_{D_i \geq 0} + \underbrace{X_i\beta + u_i}_{\equiv \varepsilon_i}. \tag{1}$$

The function $f(D)$ summarizes location-specific characteristics that affect outcomes. For example, one variable could be proximity to a city and $f(D)$ summarizes its effect on the outcome variable. More generally, $f(D)$ captures amenities and labor markets as they change across space. On the other hand, ε_i represents the *potentially* unobserved individual-specific characteristics that affect outcome variable. The quantity $Y_i(0) =$

2. See Cattaneo et al. (2019) and Cattaneo et al. (2020) for an overview of modern techniques. The formulation using first-differences is practically useful as estimation can be done using the suite of RD packages found at https://rdpackages.github.io/.

3. Keele and Titiunik (2015) discuss the choice of using a single measure of distance versus a two-dimensional running variable. The difference-in-discontinuity method can be extended into the two-dimensional framework easily, but data will usually render the two-dimensional case implausible.

$f(D_i) + \varepsilon_i$ determines the outcome variable in the absense of treatment and $\tau(D_i)$ is the average treatment effect at distance D_i. Identification of the treatment effect relies on the assumption that location-specific and individual-specific characteristics evolve smoothly across the border:

Assumption 1 (RD)**.**

(i) The functions $f(D)$ and $\tau(D)$ are continuous at the cutoff, $D = 0$,

(ii) $\mathbb{E}[\varepsilon_i \mid D_i = D]$ is continuous at the cutoff, $D = 0$.

Part (i) of assumption (RD) says that the effect of the running variable on the outcome with and without treatment is continuous at the cutoff and part (ii) says that the effect of other *potentially* unobserved individual-specific variables on the outcome is continuous at the cutoff. In the context of geographic discontinuities, a discontinuity in $f(D)$ could arise from multiple policies changing at the border and a discontinuity in $\mathbb{E}[\varepsilon_i | D_i = D]$ could arise from sorting across the border (Keele and Titiunik, 2015). These two problems represent a central threat to identification of treatment effects in the geographic RD setting.

If the two continuity assumptions are satisfied, observations in the control area close to the border identify the limiting value of $f(0)$ and observations in the treated area close to the border identify the limiting value of $f(0) + \tau(0)$. The difference between these two limits identifies $\tau(0)$. Formally, for a variable z, the left and right limits at the cutoff are denoted $z^+ \equiv \lim_{D_i \to 0^+} z(D_i)$ and $z^- \equiv \lim_{D_i \to 0^-} z(D_i)$. With assumption (RD), it is easy to show that the RD estimate identifies the treatment effect, i.e. $\tau = y^+ - y^-$.[4]

Theorem 1 (RD Identification)**.** Under assumption (RD) and model (1), $\tau(0) = y^+ - y^-$.

2.2. Difference-in-Discontinuities Identification

Now we turn to the panel setting where we observe outcomes before and after treatment occurs, $t \in \{0, 1\}$. We observe an independent and identically distributed panel sample

4. See Theorem 1 in Hahn et al. (2001).

of $\{(y_{i0}, y_{i1}, D_i)\}_{i=1}^n$. In this setting, discontinuities at the border before treatment, $t = 0$, inform us on the effects of other policy-changes and time-invariant sorting. Potential outcomes for individual i at time t are now modeled as

$$y_{it} = f_t(D_i) + \gamma(D_i)\mathbf{1}_{D_i \geq 0} + \tau(D_i)\mathbf{1}_{D_i \geq 0}\mathbf{1}_{t=1} + \varepsilon_{it}, \tag{2}$$

where $\gamma(D)$ represent the time-invariant discontinuity at the cutoff which could be due to time-invariant sorting and/or the effects of other policies that change at the border; the untreated location-specific component, $f_t(D)$, can very across periods; and $\tau(D)$ remains the treatment effect of interest.

The assumptions necessary to identify the treatment effect $\tau(0)$ requires the traditional RD assumptions to hold in both periods.

Assumption 2 (Diff-in-Disc)**.**

(i) The functions $f_0(D)$, $f_1(D)$, and $\tau(D)$ are continuous at $D = 0$,

(ii) $\mathbb{E}[\varepsilon_{it} \mid D_i = D]$ is continuous at the cutoff, $D = 0$, for $t \in \{0, 1\}$.

These assumptions warrant a bit of discussion. Note that the continuity assumption on f_0 is mild because discontinuities from other policy changes and time-invariant sorting are allowed in $\gamma(D)$. This is an improvement over traditional geographic RDs which require these effects to not be present (Keele and Titiunik, 2015).[5] In the post-period, (Diff-in-Disc) requires two things. First, $f_1(D)$ being continuous requires that no other policies turn on between periods that would cause a discontinuity at the border. Second, it requires that the effects of previous policies were already fully developed in pre-period. If the effects of other policies change over time, then the time-varying effects would not be absorbed by $\gamma(D)$ and would cause a discontinuity in $f_1(D)$ that would be mistaken as the treatment effect. Second, part (ii) requires that no additional sorting can

5. This is similar to the difference-in-differences methodology that allows for time-invariant differences in levels.

occur between 0 and 1, whether that be due to treatment or lagged sorting from previous treatments.

To help with estimation of the treatment effect, we can reformulate our potential outcomes in a first-difference model,

$$(y_1 - y_0) = (f_1(D_i) - f_0(D_i)) + \tau(D_i)\mathbf{1}_{D_i \geq 0} + (\varepsilon_{i1} - \varepsilon_{i0}),$$

where $\gamma(D_i)$ cancels out because it is time-invariant.

Theorem 2 (Diff-in-Disc Identification)**.** Under (Diff-in-Disc) and model (2), $\tau(0) = (y_1 - y_0)^+ - (y_1 - y_0)^-$.

Proof.

$$\begin{aligned}(y_1 - y_0)^+ - (y_1 - y_0)^- &= \tau(D)^+ + (f_1 - f_0)^+ + (\varepsilon_{i1} - \varepsilon_{i0})^+ - ((f_1 - f_0)^- + (\varepsilon_{i1} - \varepsilon_{i0})^-) \\ &= \tau(0) + (f_1^+ - f_1^-) + (f_0^+ - f_0^-) + (\varepsilon_1^+ - \varepsilon_1^-) + (\varepsilon_0^+ - \varepsilon_0^-) \\ &= \tau(0),\end{aligned}$$

where the second equality comes from continuity of $\tau(D)$ and the last equality comes from the two continuity assumptions. □

The above theorem says that so long as sorting and other policies are fully observed in the pre-period, a regression discontinuity estimated on a first-differenced outcome will identify the treatment effect. This theorem is closely related to Grembi et al. (2016) but differs in an important way. They find that $\tau(0) = (y_1^+ - y_1^-) - (y_0^+ - y_0^-)$ which does not require panel data. In cases of panel data, formulating the above result in terms of first-differences is advantageous. Since $(y_1 - y_0)^+ - (y_1 - y_0)^-$ is a standard RD estimate of the difference between the right and left limits, this unlocks the wide set of econometric tools used in RD estimation including local polynomial regression, data-driven bandwidth selection, and bias-corrected inference. Cattaneo et al. (2019) and Cattaneo et al. (2020) provide a literature review of the modern RD literature and include a set of `R` and `Stata` programs containing powerful estimation tools.

In the non-panel case, estimation can proceed in a local-polynomial regression framework as proposed by Grembi et al. (2016). They recommend running the following regression using observations within a small interval around $D_i = 0$:

$$\begin{aligned} Y_{it} = {} & \delta_0 + \delta_1 D_i + \mathbf{1}_{D_i \geq 0} \left(\gamma_0 + \gamma_1 D_i \right) + \\ & \mathbf{1}_{t=1} \left[\alpha_0 + \alpha_1 D_i + \mathbf{1}_{D_i \geq 0} \left(\beta_0 + \beta_1 D_i \right) \right] + \eta_{it}. \end{aligned}$$

From standard regression results, β_0, will be the difference-in-discontinuities estimate. This estimation strategy, however, does not as easily allow for the use of modern robust estimators and optimal bandwidth selection.

3 — Discussion

This paper extended the difference-in-discontinuities framework proposed by Grembi et al. (2016) into the context of geographic discontinuities. This setting faces the same problem of compound treatment that other RD contexts exhibit and since individuals can sort across the border, this context provides additional difficulties. This paper formalizes the necessary assumptions in the *geographic* context to identify the treatment effect of a policy. Moreover, in the presence of panel data, this paper proposes an improved estimation technique by recasting the estimator as a RD estimator on first-differenced data.

References

Cattaneo, Matias D., Nicolas Idrobo, and Rocio Titiunik. 2019. “A Practical Introduction to Regression Discontinuity Designs: Foundations.” *arXiv:1911.09511 [econ, stat]*. 10.1017/9781108684606, arXiv: 1911.09511.

Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2020. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Elements in Quantitative and Computational Methods for the Social Sciences, Cambridge University Press, . 10.1017/9781108684606.

Eggers, Andrew C., Ronny Freier, Veronica Grembi, and Tommaso Nannicini. 2018. “Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions: RDD Based on Population Thresholds.” *American Journal of Political Science* 62 (1): 210–229. 10.1111/ajps.12332.

Galindo-Silva, Hector, Nibene Habib Some, and Guy Tchuente. 2021. “Fuzzy Difference-in-Discontinuities: Identification Theory and Application to the Affordable Care Act.” *arXiv:1812.06537 [econ]*, `http://arxiv.org/abs/1812.06537`, arXiv: 1812.06537.

Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano. 2016. “Do Fiscal Rules Matter?” *American Economic Journal: Applied Economics* 8 (3): 1–30. 10.1257/app.20150076.

Hahn, Jinyong, Petra Todd, and Wilbert Klaauw. 2001. “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design.” *Econometrica* 69 (1): 201–209. 10.1111/1468-0262.00183.

Keele, Luke J., and Rocío Titiunik. 2015. “Geographic Boundaries as Regression Discontinuities.” *Political Analysis* 23 (1): 127–155. 10.1093/pan/mpu014.

Millán-Quijano, Jaime. 2020. “Fuzzy difference in discontinuities.” *Applied Economics Letters* 27 (19): 1552–1555. 10.1080/13504851.2019.1696930.