

Abstract

In regression discontinuity (RD), a running variable (or “score”) crossing a cutoff determines a treatment that affects the mean-regression function. This paper generalizes this usual “one-score mean RD” in three ways: (i) considering multiple scores, (ii) allowing partial effects due to each score crossing its own cutoff, not just the full effect with all scores crossing all cutoffs, and (iii) accommodating quantile/mode regressions. This generalization is motivated by (i) many multiple-score RD cases, (ii) the full-effect identification needing the partial effects to be separated, and (iii) informative quantile/mode regression functions. We establish identification for multiple-score RD (MRD), and propose simple estimators that become “local difference in differences” in case of double scores. We also provide an empirical illustration where partial effects exist.

Keywords: regression discontinuity, multiple running variables, partial effect, difference in differences

1 Introduction

Regression discontinuity (RD), originated by Thistlethwaite and Campbell (1960), has been gaining popularity in many disciplines of social sciences. Just to name a few, Rao, Yu and Ingram (2011) and Bernardi (2014) in sociology; Broockman (2009) and Caughey and Sekhon (2011) and Eggers *et al.* (2015) in political science; and many studies in economics as can be seen in the references of Imbens and Lemieux (2008), Lee and Lemieux (2010), and Choi and Lee (2017) who also list statistical papers although there are not many.

In a typical RD with a treatment D , an individual is assigned to the treatment ($D = 1$) or control group ($D = 0$), depending on a single running/forcing/assignment variable S crossing a cutoff or not. There are, however, many RD cases where multiple running variables determine a single treatment. One example is multiple test scores crossing cutoffs for school graduation or grade advancement (Jacob and Lefgren 2004). Another example is spatial/geographical RD where longitude and latitude are two running variables (Dell 2010; Keele and Titiunik 2015), although often the scalar shortest distance to a boundary is used as a running variable in the literature (Black 1999; Bayer, Ferreira and Mcmillan 2007). Since the word “running variable” will appear often in this paper, we will call it simply “score” (S for Score).

When there are multiple scores, two cases arise: “OR case” where any score can cross a cutoff to get treated (Jacob and Lefgren 2004; Matsudaira 2008; Wong, Steiner and Cook 2013), and “AND case” where all scores should cross all cutoffs to get treated. For simplification, we will examine only AND cases in this paper, because an OR case can be converted to the AND case by switching the treatment and control groups.

“Multiple-score RD (MRD) for a single treatment” that is the focus of this paper differs from “RD with multiple cutoffs for a single score” as in Angrist and Lavy (1999) and Van der Klaauw (2002),

Authors’ note: The authors are grateful to the associate editor and three anonymous reviewers for detailed comments and relevant references. The authors note that they could not accommodate all comments due to conflicts among some comments. Myoung-jae Lee’s research has been supported by a Korea University grant. Replication files for the empirical results in this paper can be found in the *Political Analysis* Dataverse (Choi and Lee 2018).

which is handled by looking at each cutoff one at a time. Whereas these studies dealt only with fixed known cutoffs, say $c_j, j = 1, \dots, J$, Cattaneo *et al.* (2016) examined a random cutoff C , which can occur in multiparty elections/races.

The goal of this paper is to generalize the usual “single-score mean-regression RD” in three ways. First, we consider multiple scores for a single treatment D . Second, differently from most other RD studies for multiple scores, we allow “partial effects” due to each score crossing its own cutoff, in addition to the (full) treatment effect due to $D = 1$ with all scores crossing all cutoffs. Third, although we focus on RD with the usual mean regression $E(Y|S)$ for a response variable Y , our approach can be easily generalized to other location measures such as conditional quantiles (Koenker 2005) and mode (Lee 1989; Kemp and Santos-Silva 2012).

Certainly, we are not the first to deal with MRD theoretically. Wong, Steiner and Cook (2013) examined “OR-case MRD,” Keele and Titiunik (2015) “AND-case MRD,” and Imbens and Zajonc (2009) and Reardon and Robinson (2012) both cases. A critical difference between these studies (except Reardon and Robinson 2012) and this paper is that we allow partial effects while they do not. To see the point, consider $S = (S_1, S_2)'$ and

$$E(Y|S) = \beta_0 + \beta_1 \delta_1 + \beta_2 \delta_2 + \beta_d D, \quad \delta_j \equiv 1[c_j \leq S_j] \quad \text{for } j = 1, 2 \text{ and } D = \delta_1 \delta_2 \quad (1)$$

where β 's are parameters, $c \equiv (c_1, c_2)'$ are known cutoffs, and $1[A] = 1$ if A holds and 0 otherwise. For instance, in the school graduation ($D = 1$) effect example (on lifetime income Y) by passing both math ($\delta_1 = 1$) and English ($\delta_2 = 1$) exams, even if one fails to have $D = 1$, still passing/failing the math exam may affect Y by encouraging/stigmatizing the student.

Ruling out partial effects, Imbens and Zajonc (2009), Wong, Steiner and Cook (2013) and Keele and Titiunik (2015) found “boundary-specific” effects, which are then weighted-averaged, in comparison to our simple effect at $S = c$ (under a weak continuity condition only at $S = c$). Reardon and Robinson (2012) seems to be the only other paper allowing for partial effects in MRD; they considered partial effects by casting MRD within a multiple treatment framework. But Reardon and Robinson (2012) did not offer formal derivations as we do in this paper.

The aforementioned generalization of $E(Y|S)$ for conditional quantiles or mode seems feasible only for “sharp RD” where D is fully determined by the scores. Hence, we stick to sharp MRD in this paper, as Wong, Steiner and Cook (2013) and Keele and Titiunik (2015) also did; only Imbens and Zajonc (2009) dealt with fuzzy MRD under no partial effects. For simplification, we will examine only two scores $S = (S_1, S_2)'$ unless otherwise noted, as generalizations to more than two scores are conceptually straightforward. Without loss of generality, we will set the cutoffs at zero unless otherwise necessary, as (S_1, S_2) can be always centered as $(S_1 - c_1, S_2 - c_2)$.

In short, we focus on AND-case two-score sharp MRD allowing partial effects for the mean-regression function $E(Y|S)$. Since the treatment D takes the interaction form $\delta_1 \delta_2$ as in Equation (1) the effect is found essentially by “local difference in differences (DD)” where both partial effects are removed in DD with only the desired interaction surviving. See Lee (2016) for the recent developments in DD.

The rest of this paper is organized as follows. Section 2 examines the identification and estimation for two-score MRD. Section 3 compares our identification conditions and estimators with those in the literature. Section 4 provides an empirical illustration. Finally, Section 5 concludes. Our MRD coverage is limited, because every issue that ever occurred to single-score RD also occurs to MRD and we cannot possibly address all the issues in one paper.

2 MRD with Two Scores

Recall $S = (S_1, S_2)'$ and $D = \delta_1 \delta_2$ where $\delta_j \equiv 1[0 \leq S_j], j = 1, 2$. First, we introduce four potential responses corresponding to $\delta_1, \delta_2 = 0, 1$, and examine partial effects—an issue that does not arise

for the usual single-score RD. Second, we impose a continuity condition and present the main identified effect for MRD. Third, we propose a simple estimation scheme based on ordinary least squares estimator (OLS) using only some observations local to the cutoff in both scores.

2.1 Four Potential Responses and Partial Effects

Define potential responses $(Y^{00}, Y^{10}, Y^{01}, Y^{11})$ corresponding to (δ_1, δ_2) being $(0, 0), (1, 0), (0, 1), (1, 1)$, respectively. Although our treatment of interest is the interaction $D = \delta_1 \delta_2$, it is possible that δ_1 and δ_2 separately affect Y . For instance, to graduate high school, one has to pass both math (δ_1) and English (δ_2) exams, but failing the math test may stigmatize the student (“I cannot do math”) to affect his/her lifetime income Y ; in this case, Y is affected by δ_1 as well as by D . More generally, when an interaction term appears in a regression function, it is natural to allow the individual terms in the regression function. Call the separate effects of δ_1 and δ_2 “partial effects.”

At a glance, the individual treatment effect of interest may look like $Y^{11} - Y^{00}$ because $D = \delta_1 \delta_2$, but this is not the case. To see why, think of the high school graduation example. Y^{11} is the lifetime income when both exams are passed, and as such, Y^{11} includes the high school graduation effect on lifetime income and the partial effect of passing the math exam (“I can do math”), as well as the possible partial effect of passing the English exam (“I can do English?”). Hence the “net” effect of high school graduation should be

$$Y^{11} - Y^{00} - (Y^{10} - Y^{00}) - (Y^{01} - Y^{00}) = Y^{11} - Y^{10} - Y^{01} + Y^{00}$$

where the two partial effects relative to Y^{00} are subtracted from $Y^{11} - Y^{00}$.

Rewrite $E(Y|S)$ as

$$E(Y|S) = E(Y^{00}|S)(1 - \delta_1)(1 - \delta_2) + E(Y^{10}|S)\delta_1(1 - \delta_2) + E(Y^{01}|S)(1 - \delta_1)\delta_2 + E(Y^{11}|S)\delta_1\delta_2. \tag{2}$$

Further rewrite this so that δ_1 and δ_2 and $D = \delta_1 \delta_2$ appear separately:

$$E(Y|S) = E(Y^{00}|S) + \{E(Y^{10}|S) - E(Y^{00}|S)\}\delta_1 + \{E(Y^{01}|S) - E(Y^{00}|S)\}\delta_2 + \{E(Y^{11}|S) - E(Y^{10}|S) - E(Y^{01}|S) + E(Y^{00}|S)\}D \tag{3}$$

which will play the main role for MRD. This equation does not hold for fuzzy RD, because D would then depend on random variables other than S on the right-hand side while the left-hand side $E(Y|S)$ is a function of only S . This is one of the reasons why we stick to sharp RD.

The slope of $D = \delta_1 \delta_2$ in Equation (3) is reminiscent of the above $Y^{11} - Y^{10} - Y^{01} + Y^{00}$, and it is a DD with $E(Y^{11}|S) - E(Y^{10}|S)$ as the “treatment group difference” and $E(Y^{01}|S) - E(Y^{00}|S)$ as the “control group difference.” Since D is an interaction, it is only natural that DD is used to find the treatment effect, as DD is known to isolate the interaction effect by removing the partial effects.

If

$$\text{no partial effects : } E(Y^{10}|S) = E(Y^{01}|S) = E(Y^{00}|S),$$

then Equation (3) becomes

$$E(Y|S) = E(Y^{00}|S) + \{E(Y^{11}|S) - E(Y^{00}|S)\}D. \tag{4}$$

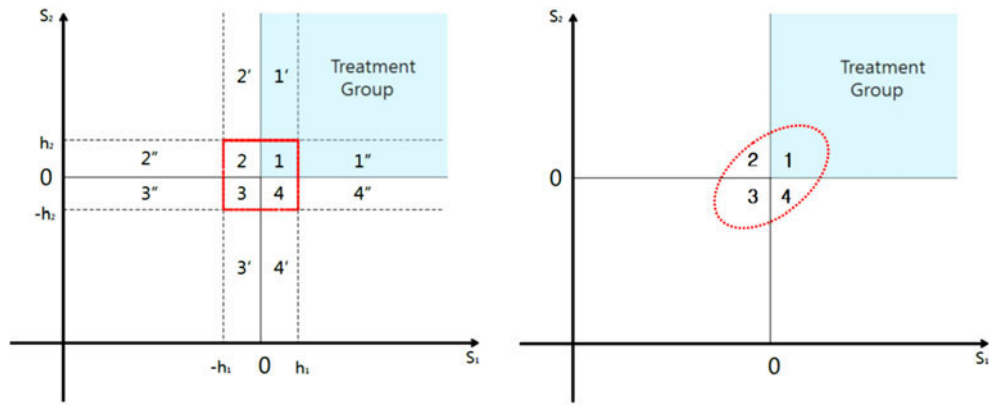


Figure 1. Two-Score RD in AND case (Square & Oval Neighborhoods).

It helps to see when the no partial-effect assumption is violated (recall Equation (1) with $\beta_1 \neq 0$ or $\beta_2 \neq 0$):

$$E(Y^{11}) = \beta_0 + \beta_1 + \beta_2 + \beta_d, \quad E(Y^{10}) = \beta_0 + \beta_1, \quad E(Y^{01}) = \beta_0 + \beta_2, \quad E(Y^{00}) = \beta_0$$

$$\implies E(Y^{11}) - E(Y^{10}) - E(Y^{01}) + E(Y^{00}) = \beta_d, \quad E(Y^{11}) - E(Y^{00}) = \beta_1 + \beta_2 + \beta_d.$$

Examine squares 1–4 in the left panel of Figure 1, where (h_1, h_2) are the two localizing bandwidths. There is one treatment group (square 1) and three control groups (squares 2, 3 and 4). Under no partial effect, the treatment effect can be found by comparing squares 1 and 2, 1 and 4, or 1 and 3. With partial effects present, however, this is no longer the case: squares 1 and 2 give the treatment effect β_d plus the partial effect due to S_1 crossing 0; squares 1 and 4 give β_d plus the partial effect due to S_2 crossing 0; squares 1 and 3 give β_d plus the two partial effects. It is only when we take DD as in Equation (3) that the desired β_d is identified. More generally than the left panel of Figure 1, we may have the right panel where the four groups are not squares, but parts of an oval figure depending on the correlation between S_1 and S_2 .

2.2 Identification and Remarks

To simplify notation for limits of $E(Y|S = s) = E(Y|S_1 = s_1, S_2 = s_2)$, denote

$$\lim_{s_1 \downarrow 0, s_2 \downarrow 0} \text{ as } \lim_{+,+}, \quad \lim_{s_1 \uparrow 0, s_2 \downarrow 0} \text{ as } \lim_{-,+}, \quad \lim_{s_1 \downarrow 0, s_2 \uparrow 0} \text{ as } \lim_{+,-}, \quad \lim_{s_1 \uparrow 0, s_2 \uparrow 0} \text{ as } \lim_{-,-}.$$

Assume that these double limits of $E(\cdot|S)$ exist at 0 for the potential responses, and denote them using 0^- and 0^+ ; for example, $E(Y^{00}|0^-, 0^+) \equiv \lim_{-,+} E(Y^{00}|s_1, s_2)$.

Take the double limits on Equation (2) to get

$$E(Y|0^+, 0^+) = E(Y^{11}|0^+, 0^+), \quad E(Y|0^+, 0^-) = E(Y^{10}|0^+, 0^-),$$

$$E(Y|0^-, 0^+) = E(Y^{01}|0^-, 0^+), \quad E(Y|0^-, 0^-) = E(Y^{00}|0^-, 0^-). \tag{5}$$

These give a limiting version of the slope of $D = \delta_1 \delta_2$ in Equation (3) at $(0, 0)$:

$$E(Y|0^+, 0^+) - E(Y|0^+, 0^-) - E(Y|0^-, 0^+) + E(Y|0^-, 0^-)$$

$$= E(Y^{11}|0^+, 0^+) - E(Y^{10}|0^+, 0^-) - E(Y^{01}|0^-, 0^+) + E(Y^{00}|0^-, 0^-). \tag{6}$$

Assume the continuity condition (note that all right-hand side terms have $(0^+, 0^+)$)

$$\begin{aligned}
 \text{(i): } & E(Y^{01}|0^-, 0^+) = E(Y^{01}|0^+, 0^+), \\
 \text{(ii): } & E(Y^{10}|0^+, 0^-) = E(Y^{10}|0^+, 0^+), \\
 \text{(iii): } & E(Y^{00}|0^-, 0^-) = E(Y^{00}|0^+, 0^+).
 \end{aligned}
 \tag{7}$$

Equation (7)(i) is plausible because Y^{01} is untreated along s_1 , (ii) because Y^{10} is untreated along s_2 , and (iii) because Y^{00} is untreated along both s_1 and s_2 . These continuity conditions show how counterfactuals for the treatment group with $(0^+, 0^+)$ can be identified. For example, Equation (7)(i) is that the counterfactual $E(Y^{01}|0^+, 0^+)$ for the treatment group can be identified with $E(Y^{01}|0^-, 0^+)$ from the partially treated group $(0^-, 0^+)$.

Using Equations (7), (6) becomes

$$E(Y|0^+, 0^+) - E(Y|0^+, 0^-) - E(Y|0^-, 0^+) + E(Y|0^-, 0^-)
 \tag{8}$$

$$\begin{aligned}
 &= \beta_d \equiv E(Y^{11}|0^+, 0^+) - E(Y^{10}|0^+, 0^+) - E(Y^{01}|0^+, 0^+) + E(Y^{00}|0^+, 0^+) \\
 &= E(Y^{11} - Y^{10} - Y^{01} + Y^{00}|0^+, 0^+);
 \end{aligned}
 \tag{9}$$

Equation (8) is an identified entity that is characterized by Equation (9)—the mean effect on the just treated $(0^+, 0^+)$. We summarize this (as well as Equation (4) under no partial effect) as a theorem, with a three-score MRD extension provided in the appendix A.

THEOREM 1. *Suppose the double limits of $E(Y|S)$ exist at 0 for the potential responses, the continuity condition Equation (7) holds, and the density function $f_S(s)$ of S is strictly positive on a neighborhood of $(0, 0)$. Then the effect*

$$\beta_d = E(Y^{11} - Y^{10} - Y^{01} + Y^{00}|0^+, 0^+)$$

is identified by two-score MRD Equation (8). If no partial-effect condition holds at $S = 0$ (i.e., $E(Y^{10}|0^+, 0^+) = E(Y^{01}|0^+, 0^+) = E(Y^{00}|0^+, 0^+)$), then $\beta_d = E(Y^{11} - Y^{00}|0^+, 0^+)$.

Would partial effects really matter? Partial effects may be unlikely in certain MRDs. For instance, in two-dimensional geographic MRD with latitude S_1 and longitude S_2 , simply crossing only one boundary may not do much of anything. But if $S_2 \geq 0$ corresponds to being on the right side of mountains ranging south to north, then a partial effect due to S_2 can occur, because the weather on the right side of the mountain range can be much different from that on the left side. Another example is the effects of a conservative party being the majority in both houses of parliament on the passage of bills, where the cutoff is 50% of the seats in each house. Even if the conservative party is the majority in only one of the two houses, still the passage rate can be different from when the conservative party is not the majority in either house. Given that allowing for partial effects is not difficult at all as can be seen shortly, there is no reason to simply assume away partial effects.

2.3 OLS

Although Equation (8) shows that β_d can be estimated by replacing the four identified elements in Equation (8) with their sample versions, in practice, it is easier to implement MRD with Equation (3), using only the local observations satisfying $S_j \in (-h_j, h_j)$, $j = 1, 2$. Specifically, replace $E(Y^{00}|S)$ in Equation (3) with a (piecewise-) continuous function of S , and replace the slopes of δ_1 , δ_2 and D with parameters β_1 , β_2 and β_d to obtain

$$E(Y|S) = E(Y^{00}|S) + \beta_1\delta_1 + \beta_2\delta_2 + \beta_d D
 \tag{10}$$

where $E(Y^{00}|S)$ is specified as

$$\begin{aligned} \text{linear: } m_1(S) &\equiv \text{a linear function of } S_1, S_2 \text{ with intercept } \beta_0 \\ \text{quadratic: } m_2(S) &\equiv m_1(S) + \text{a linear function of } S_1^2, S_2^2, S_1 S_2. \end{aligned} \tag{11}$$

Then OLS can be applied to Equation (10) to do inference with the usual OLS asymptotic variance estimator. If $E(\cdot|S)$ in Equation (10) is replaced with a conditional quantile/mode, quantile/mode regression can be applied to estimate the quantile/modal parameters.

With

$$\delta_j^- \equiv 1[-h_j < S_j < 0], \quad \delta_j^+ \equiv 1[0 \leq S_j < h_j], \quad j = 1, 2,$$

another way to set $E(Y^{00}|S)$ is a piecewise-linear function continuous at 0:

$$\begin{aligned} E(Y^{00}|S) &= \beta_0 + \beta_{11} \delta_1^- \delta_2^- S_1 + \beta_{12} \delta_1^- \delta_2^- S_2 + \beta_{21} \delta_1^- \delta_2^+ S_1 + \beta_{22} \delta_1^- \delta_2^+ S_2 \\ &+ \beta_{31} \delta_1^+ \delta_2^- S_1 + \beta_{32} \delta_1^+ \delta_2^- S_2 + \beta_{41} \delta_1^+ \delta_2^+ S_1 + \beta_{42} \delta_1^+ \delta_2^+ S_2. \end{aligned} \tag{12}$$

This allows different slopes across the four quadrants determined by $(\delta_1^-, \delta_1^+, \delta_2^-, \delta_2^+)$.

The above MRD estimation requires choosing the functional form for $E(Y^{00}|S)$, $h \equiv (h_1, h_2)'$ for S , and a weighting function within the chosen local neighborhood. First, we use only a linear or quadratic function of S in Equations (11) and (12), as Gelman and Imbens (2018) advise against using high-order polynomials in RD. Second, developing optimal bandwidths for h in MRD as Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014) did for single-score RD would be very involved, going over the scope of this paper; instead, we use a rule-of-thumb bandwidth $N^{-1/6}$ with both scores standardized, and explore cross validation (CV) schemes below to find useful reference bandwidths. Third, we do not use any weighting function within the chosen local neighborhood in the above OLS, which amounts to adopting the uniform weight; this is a common practice, as weighting seems to make little difference in practice. There is no proof that these choices that we make are optimal, which means that our proposed estimation strategy in this section to be applied in the empirical section should be taken as tentative; hopefully, further research settles the estimation issues in a more satisfactory manner.

In RD, the sample size can be small due to the localization, and the problem gets exacerbated for MRD. In case this happens, Cattaneo, Frandsen and Titiunik (2015), Keele, Titiunik and Zubizarreta (2015) and Cattaneo, Titiunik and Vazquez-Bare (2017) proposed “randomized inference.” But applying this to MRD is challenging, because randomly assigning each subject to one of the four groups under the null of no effect requires the null hypothesis to be $\beta_1 = \beta_2 = \beta_d = 0$ in Equation (10) instead of only $\beta_d = 0$ while allowing $\beta_1 \neq 0$ or $\beta_2 \neq 0$, which was the very motivation for this paper. Designing a proper randomized inference for MRD is an interesting research question, but it goes beyond the scope of this paper.

About choosing h , one CV scheme for MRD is minimizing

$$\sum_i \omega_i^h \{Y_i - \tilde{E}_{-i}(Y|S_i, h)\}^2 / \sum_i \omega_i^h \tag{13}$$

with respect to h , where $\omega_i^h = 1$ for S_i with at least 2 or 3 observations in each of the four directions within its “square neighborhood” $(S_{1i} \pm h_1, S_{2i} \pm h_2)$, and $\omega_i^h = 0$ otherwise; this ensures ruling out S_i ’s on its support boundaries. In this CV scheme, $\tilde{E}_{-i}(Y|S_i, h)$ is a nonparametric kernel predictor using an one-sided kernel estimator depending on the side of $(0, 0)$ where S_i is located among the four sides, which is a generalization of the CV scheme in Ludwig and Miller (2007) who applied Equation (13) to single-score RD. As it turned out, however, we experienced the same problem as

Ludwig and Miller (2007) experienced: too large bandwidths that make most ω_i^h 's zero and predict the few remaining Y_i 's well to make Equation (13) small.

The problem of too large bandwidths does not occur to the “conventional CV” which uses all-sided symmetric weighting to minimize

$$\frac{1}{N} \sum_i \{Y_i - \hat{E}_{-i}(Y|S_i, h)\}^2 \quad \text{where } \hat{E}_{-i}(Y|S_i, h) \equiv \frac{\sum_{j \neq i} K_h(S_j - S_i) Y_j}{\sum_{j \neq i} K_h(S_j - S_i)} \quad (14)$$

and K_h is a kernel function with bandwidths h . This is known to behave well: the resulting minimand is nearly convex and the conventional CV bandwidth is asymptotically optimal. The reason why this is not used in single-score RD is that $E(Y|S)$ has a break, instead of being continuous in S , and consequently $\hat{E}_{-i}(Y|S_i, h)$ is biased for $E(Y|S_i)$ when S_i is near the cutoff. Nevertheless, since the goal is finding a reasonable h , not necessarily predicting Y well, we use this conventional CV.

Although we adopt the uniform weight within a chosen neighborhood, still the neighborhood should be chosen whose form differs as Figure 1 illustrates. With $\rho \equiv \text{COR}(S_1, S_2)$, $\sigma_j \equiv \text{SD}(S_j)$ and $\eta_j \equiv h_j/\sigma_j$ ($\iff h_j \equiv \sigma_j \eta_j$) for $j = 1, 2$, we use

$$\begin{aligned} \text{(i) square-neighbor kernel: } K_h(S) &= 1 \left[\left| \frac{S_1}{\sigma_1 \eta_1} \right| \leq 1 \right] \cdot 1 \left[\left| \frac{S_2}{\sigma_2 \eta_2} \right| \leq 1 \right], \\ \text{(ii) oval-neighbor kernel: } K_h(S) &= 1 \left[\left(\frac{S_1}{\sigma_1 \eta_1} \right)^2 - 2\rho \frac{S_1}{\sigma_1 \eta_1} \frac{S_2}{\sigma_2 \eta_2} + \left(\frac{S_2}{\sigma_2 \eta_2} \right)^2 \leq 1 \right]. \end{aligned} \quad (15)$$

These kernels need normalizing factors, but they are irrelevant in choosing η_1 and η_2 because they get canceled in $\hat{E}_{-i}(Y|S_i, h)$.

Setting $\eta_1 = \eta_2 \equiv \eta$ in Equation (15)(i) gives a square neighborhood of 0 in the standardized scores ($S_1/\sigma_1, S_2/\sigma_2$) and setting $\eta_1 = \eta_2 \equiv \eta$ and $\rho = 0$ in Equation (15)(ii) gives a circle because the two kernels become

$$1 \left[\left| \frac{S_1}{\sigma_1} \right| \leq \eta \right] \cdot 1 \left[\left| \frac{S_2}{\sigma_2} \right| \leq \eta \right] \quad \text{and} \quad 1 \left[\left(\frac{S_1}{\sigma_1} \right)^2 + \left(\frac{S_2}{\sigma_2} \right)^2 \leq \eta^2 \right].$$

The oval shape is elongated along the 45 degree line when $\rho > 0$ as in the right panel of Figure 1, and such a neighborhood can better capture observations scattered along the 45 degree line; when $\rho < 0$, the oval shape is elongated along the 135 degree line.

3 Other Approaches in the Literature

Having presented our proposal, now we review the other approaches for MRD. First, two scores are collapsed into one so that the familiar single-score RD arsenal can be mobilized. Second, two-dimensional localization is avoided by doing, for example, one-dimensional localization for S_1 given $S_2 \geq 0$ (i.e., given $\delta_2 = 1$) to get the “effects on the boundary $S_1 = 0$ ”; here as well, the familiar single-score RD methods can be utilized. Third, those effects on the boundary can be weight-averaged.

3.1 Minimum Score

Battistin *et al.* (2009) and Clark and Martorell (2014) defined

$$S_m \equiv \min(S_1, S_2) \implies D = 1[0 \leq S_m]$$

to set up

$$E(Y|S_m) = \beta_0 + \beta_- S_m(1 - D) + \beta_+ S_m D + \beta_m D$$

where β_m is the treatment effect of interest. Recalling Equation (10) with $\beta_1 = \beta_2 = 0$, we can see that $E(Y^{00}|S_1, S_2)$ in Equation (10) is specified just as $\beta_0 + \beta_- S_m(1 - D) + \beta_+ S_m D$.

This approach is problematic because the linear spline $\beta_0 + \beta_- S_m(1 - D) + \beta_+ S_m D$ is inadequate: it approximates $E(Y^{00}|S)$ only with S_1 when $S_1 < S_2$, and only with S_2 when $S_2 < S_1$ —there is no reason to voluntarily “handcuff” oneself this way, and better approximations can be seen in Equations (11) and (12). Also, partial effects are ruled out because $\beta_0 + \beta_- S_m(1 - D) + \beta_+ S_m D$ is continuous in S_m that is in turn continuous in S : no break along S_1 only (nor S_2 only) is allowed.

A couple of remarks are in order. First, Reardon and Robinson (2012) and Wong, Steiner and Cook (2013) called this approach, respectively, “binding score approach” and “centering approach,” but “min approach” would be more fitting. Second, Battistin *et al.* (2009) and Clark and Martorell (2014) dealt with fuzzy mean-based MRDs, not sharp MRD. Third, S_m can be easily generalized to more than two scores; for example, $\min(S_1, S_2, S_3)$ for three scores as in Clark and Martorell (2014).

3.2 One-Dimensional Localization

The dominant approach in the MRD literature is looking at a subpopulation with one score already greater than its cutoff (Jacob and Lefgren 2004; Lalive 2008; Matsudaira 2008). For instance, on the subpopulation with $\delta_1 = 1$, δ_2 equals D , and squares 1 and 1'' in the left panel of Figure 1 become the treatment group whereas squares 4 and 4'' become the control group. This raises efficiency because only one-dimensional localization is done with the larger control and treatment groups, but a bias appears if there is a partial effect. Reardon and Robinson (2012) and Wong, Steiner and Cook (2013) called this “frontier approach” and “univariate approach,” respectively.

To formalize the idea, set $\delta_1 = 1 (\iff S_1 \geq 0)$ and $D = \delta_2$ in Equation (3) to have

$$E(Y|S) = E(Y^{10}|S) + \{E(Y^{11}|S) - E(Y^{10}|S)\}\delta_2; \tag{16}$$

$E(Y^{10}|S)$ is the baseline now. Take the upper and lower limits only for s_2 with $s_1 \geq 0$:

$$\begin{aligned} E(Y|s_1, 0^+) &= E(Y^{10}|s_1, 0^+) + \lim_{s_2 \downarrow 0} \{E(Y^{11}|s_1, s_2) - E(Y^{10}|s_1, s_2)\}, \\ E(Y|s_1, 0^-) &= E(Y^{10}|s_1, 0^-). \end{aligned}$$

Assume the continuity condition

$$E(Y^{10}|s_1, 0^+) = E(Y^{10}|s_1, 0^-) \quad \forall s_1 \geq 0; \tag{17}$$

whereas this has “ $\forall s_1 \geq 0$,” (ii) of Equation (7) is only for $s_1 = 0^+$ that is weaker than Equation (17). Using Equation (17), the difference between the upper and lower limits gives

$$\beta^{10}(s_1, 0^+) \equiv \lim_{s_2 \downarrow 0} \{E(Y^{11}|s_1, s_2) - E(Y^{10}|s_1, s_2)\} = E(Y|s_1, 0^+) - E(Y|s_1, 0^-);$$

“10” in $\beta^{10}(s_1, 0^+)$ refers to the baseline superscript in Y^{10} . For Equation (1), $\beta^{10}(s_1, 0^+) = \beta_2 + \beta_d$, not β_d .

Proceeding analogously, set $\delta_2 = 1 (\iff S_2 \geq 0)$ and $D = \delta_1$ in Equation (3) to have

$$E(Y|S) = E(Y^{01}|S) + \{E(Y^{11}|S) - E(Y^{01}|S)\}\delta_1. \tag{18}$$

Take the upper and lower limits only for s_1 with $s_2 \geq 0$:

$$E(Y|0^+, s_2) = E(Y^{01}|0^+, s_2) + \lim_{s_1 \downarrow 0} \{E(Y^{11}|s_1, s_2) - E(Y^{01}|s_1, s_2)\},$$

$$E(Y|0^-, s_2) = E(Y^{01}|0^-, s_2).$$

Assume the continuity condition

$$E(Y^{01}|0^+, s_2) = E(Y^{01}|0^-, s_2) \quad \forall s_2 \geq 0. \tag{19}$$

Using Equation (19), the difference between the upper and lower limits gives

$$\beta^{01}(0^+, s_2) \equiv \lim_{s_1 \downarrow 0} \{E(Y^{11}|s_1, s_2) - E(Y^{01}|s_1, s_2)\} = E(Y|0^+, s_2) - E(Y|0^-, s_2).$$

For Equation (1), $\beta^{01}(0^+, s_2) = \beta_1 + \beta_d$, not β_d .

In estimation for Equation (16), the usual single-score RD approach would adopt

$$E(Y|S) = E(Y^{10}|S) + \beta^{10} \delta_2 \quad \text{for a parameter } \beta^{10} \tag{20}$$

analogously to Equation (10), where $E(Y^{10}|S)$ is specified as in Equation (11); only the subsample with $(\delta_2^- + \delta_2^+) \delta_1 = 1$ is used for estimation. There is no “oval-neighbor” analog, because only the observations with $|S_2| \leq h_2$ are used given $S_1 \geq 0$.

The model Equation (20) may be inadequate, because S_1 in the slope of δ_2 in Equation (16) is not localized. That is, replacing β^{10} in Equation (20) with a function of S_1 would be better, which then results in a model such as

$$E(Y|S) = E(Y^{10}|S) + \beta^{12} S_1 \delta_2 + \beta^{10} \delta_2 \quad \text{for a parameter } \beta^{12}. \tag{21}$$

For the opposite case of localizing with S_1 given $S_2 \geq 0$, we can use analogously

$$E(Y|S) = E(Y^{01}|S) + \beta^{01} \delta_1 \quad \text{or} \quad E(Y|S) = E(Y^{01}|S) + \beta^{21} S_2 \delta_1 + \beta^{01} \delta_1.$$

3.3 Weighted Average of Boundary Effects

Imbens and Zajonc (2009) dealt with both multiple-score sharp RD and fuzzy RD in a general set-up allowing both AND and OR cases. They discussed identification and estimation, assuming away partial effects. With B denoting the treatment and control boundary, the treatment effect at $s \in B$ for FRD is

$$\beta_d(s) \equiv \frac{\lim_{v \rightarrow 0} E\{Y|S \in N_v^+(s)\} - \lim_{v \rightarrow 0} E\{Y|S \in N_v^-(s)\}}{\lim_{v \rightarrow 0} E\{D|S \in N_v^+(s)\} - \lim_{v \rightarrow 0} E\{D|S \in N_v^-(s)\}}$$

where $N_v^+(s)$ and $N_v^-(s)$ denote the “ v -treated”- and “ v -control” neighborhoods of s .

Imbens and Zajonc (2009) proposed also an integrated version of $\beta_d(s)$:

$$\beta_d \equiv \int_{s \in B} \beta_d(s) f_S(s|S \in B) \partial s = \frac{\int_{s \in B} \beta_d(s) f_S(s) \partial s}{\int_{s \in B} f_S(s) \partial s}. \tag{22}$$

Tests for the effect heterogeneity along B and the asymptotic distribution using a multivariate local linear regression are also shown in Imbens and Zajonc (2009).

Keele and Titiunik (2015; “KT”) addressed AND-case two-score sharp MRD. Consider the two boundary lines B stemming from the cutoff (c_1, c_2) rightward and upward as in the left panel of

Figure 1. With partial effects ruled out in KT, only the treatment gets administered as B is crossed to the “treatment quadrant” ($c_1 \leq S_1, c_2 \leq S_2$) from any direction. Denoting a point in B as b , KT assumed the continuity at all points in B for the potential untreated and treated responses Y_0 and Y_1 :

$$\lim_{s \rightarrow b} E(Y_0|S = s) = E(Y_0|S = b) \quad \text{and} \quad \lim_{s \rightarrow b} E(Y_1|S = s) = E(Y_1|S = b).$$

Denoting a point in the treatment quadrant as s^t and in the control quadrants as s^c , this continuity condition identifies the effect $\tau(b)$ at $b \in B$:

$$\begin{aligned} \lim_{s^t \rightarrow b} E(Y|S = s^t) - \lim_{s^c \rightarrow b} E(Y|S = s^c) &= \lim_{s^t \rightarrow b} E(Y_1|S = s^t) - \lim_{s^c \rightarrow b} E(Y_0|S = s^c) \\ &= E(Y_1|S = b) - E(Y_0|S = b) = E(Y_1 - Y_0|S = b) \equiv \tau(b). \end{aligned} \tag{23}$$

A marginal effect can be found by integrating out b as in Equation (22). KT proposed a local polynomial regression estimator for $\tau(b)$ using a distance from b , say the Euclidean distance $\lambda_b(S) \equiv \|S - b\|$, as a single “regressor.” This is to be done on the treatment and control quadrants separately to obtain sample analogs for the first term of Equation (23). The difference of the intercept estimators is then an estimator for $\tau(b)$.

Wong, Steiner and Cook (2013; “WSC”) dealt with OR-case two-score sharp MRD where $D = 1[S_1 < c_1 \text{ or } S_2 < c_2]$; WSC ruled out partial effects. WSC laid out four approaches, and we explain three (the remaining one does not seem tenable, and WSC did not recommend it either). The first is the aforementioned minimum of the scores. The second is essentially the one-dimensional localization along the horizontal boundary (say B_1) of B , and then along the vertical boundary (say B_2); the difference from KT is, however, that WSC obtained $\tau_1 \equiv E(Y_1 - Y_0|S \in B_1)$ and $\tau_2 \equiv E(Y_1 - Y_0|S \in B_2)$ instead of KT’s $E(Y_1 - Y_0|S = b)$ for all $b \in B$. The third is getting an weighted average of τ_1 and τ_2 , which WSC called the “frontier average treatment effect.”

Although disallowing partial effects may look simplifying, to the contrary, it results in considering boundary lines instead of the single boundary point (c_1, c_2). The possibly heterogeneous effects along the boundaries may be informative, and possibly efficiency enhancing if they are homogeneous, which however also raises the issue of finding a single marginal effect as a weighted average of those boundary effects. Such a weighting requires estimating densities for the boundary lines—a complicating scenario.

Of course, in reality, whether partial effects exist or not is an empirical question. The logical thing to do is thus to allow nonzero partial effects first with our approach, and then test for zero partial effects; if accepted, one may adopt some of the above approaches. This should be preferred than simply ruling out partial effects from the beginning, unless there is a strong prior justification to do so.

4 Empirical Illustration

This section provides an empirical example for congress “productivity”: the effects of the Republican party being dominant in both lower and upper houses on passing bills, where the sample size is only 104. We estimate the mean effect, but the inference is problematic due to the small sample size. Also, the usual RD data plots are not helpful, because dividing the range of S to create cells leaves only a few observations for each cell. We use two measures of legislative productivity for the US Congress 1789–2004 in Grant and Kelly (2008): the “legislative productivity index (LPI)” for all legislations, and the “major legislation index (MLI)” for major legislations only. We obtained the House (S_1) and Senate (S_2) Republican seat proportions from

Table 1. Descriptive Statistics for Congress Productivity Data.

	All Sample		± 0.10	± 0.05
	Mean (SD)	Min, Max	Mean (SD)	Mean (SD)
S_1 (House)	0.498 (0.139)	0.202, 0.860	0.494 (0.057)	0.497 (0.028)
S_2 (Senate)	0.517 (0.147)	0.167, 0.917	0.504 (0.048)	0.502 (0.019)
D	0.404 (0.493)	0, 1	0.405 (0.497)	0.357 (0.497)
LPI	90.8 (57.7)	3.9, 187	98.5 (55.5)	108 (50.8)
MLI	11.1 (5.20)	3.1, 20.3	11.7 (5.24)	12.1 (4.60)
N	104		42	14

<http://www.senate.gov/history/partydiv.htm> and <http://history.house.gov/Institution/Party-Divisions/Party-Divisions/>.¹

For the periods before 1837, we consider Jackson, Jackson Republican, Jeffersonian Republican, and Anti-Administration as Republican parties to follow the party division that the official Senate and House website makes. Since there was no official Republican party before 1857, for 1837–1856, we consider the parties opposite to the Democratic party as Republican. Among the total 108 congresses, we removed four cases where neither Democrats nor Republicans were dominant.

Table 1 presents descriptive statistics. On average, the Republican seat proportions are around 0.5 and they are the majority in both houses 40% of the times. LPI is 90.8 on average and MLI is 11.1, and when we restrict the sample to ± 0.10 around the cutoff 0.5 in both houses to have 42 observations, the average LPI increases to 98.5 and MLI to 11.7. When we restrict the sample to ± 0.05 , the average LPI further increases to 108 and MLI to 12.1—but then only 14 observations are left.

Figure 2 plots LPI and MLI, which reveals an increasing trend. We do the OLS of Y on $(1, t, S_1, S_2, \delta_1, \delta_2, D)$, where t is to capture the trend, Y is standardized to ease interpretation (i.e., Y is $LPI/SD(LPI)$ or $MLI/SD(MLI)$), and D is the indicator for whether the Republican party is dominant in both houses or not; other than t , we adopt Equation (10) with the linear model in Equation (11). We also tried using t^2 additionally, but the results are omitted as they do not differ much.

Although there is no covariate in our data, the lagged outcome can be thought of as a covariate, which may be unbalanced between the treatment and control groups. To check this out, we do the OLS of the lagged Y on the same regressors to test $H_0 : \beta_1 = \beta_2 = \beta_d = 0$ (i.e., balance across the treatment and three control groups in the lagged Y). For three bandwidths 0.05, 0.10, 0.15 with both scores standardized, the p values of the test are

LPI : 0.008, 0.210, 0.104

MLI : 0.245, 0.133, 0.469

The test rejects for LPI with bandwidth 0.05, which may very well be due to the small sample size 14, because 14 means 3.7 observations per group for which law of large numbers can hardly work. For the other cases, the test does not reject.

For $h = (h_1, h_2)'$, we use the single rule-of-thumb bandwidth $SD(S_j)N^{-1/6} = 0.065$ for $j = 1, 2$ due to $SD(S_1) \approx SD(S_2) \approx 0.14$ in Table 1, and the CV bandwidths described in Equation (14). For CV, we try a common single bandwidth ($\eta_c \equiv \eta_1 = \eta_2$) or two different bandwidths $\eta_1 \neq \eta_2$ using the square or oval-neighbor kernels in Equation (15). For the common single bandwidth, the CV gave $SD(S_j)\eta_c = 0.09$ with the square-neighbor kernel, and 0.12 with the oval-neighborhood

1 Replication files for the empirical results in this paper can be found in the *Political Analysis* Dataverse (Choi and Lee 2018).

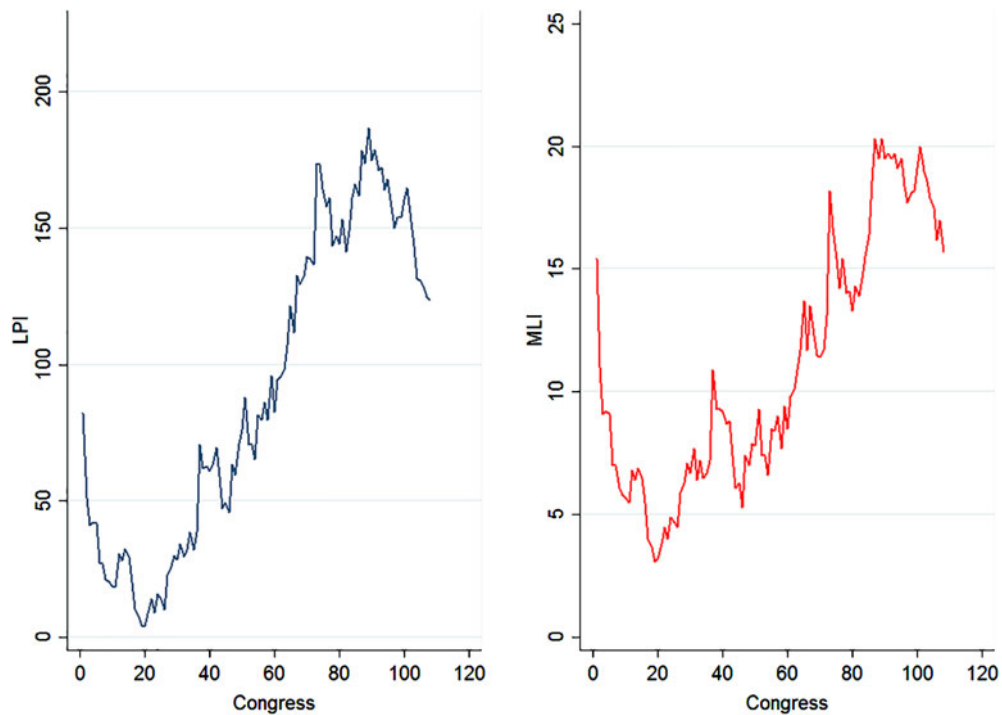


Figure 2. LPI and MLI across Congresses.

kernel. When we allowed $\eta_1 \neq \eta_2$, the square-neighbor kernel gave $(h_1, h_2) = (0.07, 0.12)$, and the oval-neighbor kernel gave $(h_1, h_2) = (0.10, 0.12)$. The local observations selected by these four different bandwidths are shown in Figure 3; since $\text{COR}(S_1, S_2) = 0.76$ in our data, the observations are scattered along the 45 degree line with most observations in quadrants 1 and 3. Overall, the CV bandwidths range over 0.07 to 0.12, and the rule-thumb bandwidth 0.065 is almost the same as the smallest CV bandwidth 0.07.

The estimation results for LPI and MLI are in Tables 2 and 3, each with three panels. In the first panel, “Sq” stands for square-neighbor kernel, “RT” stands for rule-of-thumb bandwidth, CV1 is CV with one common bandwidth, and CV2 is CV with two bandwidths. The row “ N_1-N_4 ” lists the local number of observations in the four quadrants, and the row “ $\sum_j N_j/N$ ” shows the proportion of the used local observations relative to the total number of available observations $N = 104$. The second panel shows the treatment effect estimates by our proposal (OLS) and the existing methods in the literature: BW for the boundary-estimate-weighting method in Equation (22), MIN for $\min(S_1, S_2)$, RD1 for one-dimensional RD with $S_1|S_2 = 1$ in (3.3), and RD2 for one-dimensional RD with $S_2|S_1 = 1$ in Equation (16). The third panel presents the partial- effect estimates by our proposed OLS.

BW did not work with the rule-of-thumb bandwidth because it is too small to have enough observations in each neighbor of all boundary points. Since MIN, RD1 and RD2 use unidimensional “square” neighbor, we put their estimates in the “Sq” columns. For inference, 90% and 95% confidence intervals (CI) were calculated from bootstrap with 10,000 repetitions because the sample size is small. The statistical significance is determined by whether the CI captures zero or not; to save space, we present only 95% CIs.

In Table 2 for LPI, Oval-CV1, and Oval-CV2 use more than 50% of the available data, going away from RD localization; hence we would trust the other columns (Sq-RT, Sq-CV, and Oval-RT) more, where the treatment effects fall in 0.62–1.39 which are statistically significant. These numbers differ much from the estimates from the existing methods in the literature. This difference is understandable in view of the significant partial effect β_2 ranging over -0.38 – 0.72 in the columns

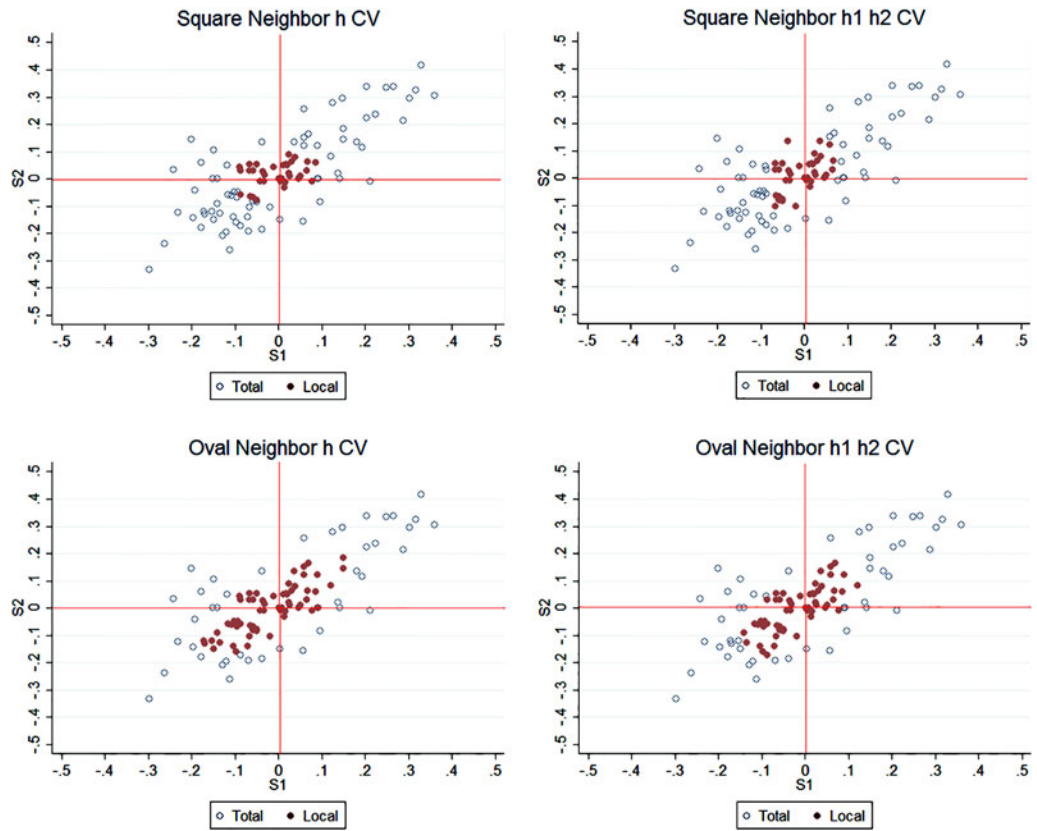


Figure 3. Square & Oval Neighbors (1 & 2 Bandwidths) Choose Different Observations.

Table 2. LPI Estimates for Treatment and Partial Effects.

	Sq-RT	Sq-CV1	Sq-CV2	Oval-RT	Oval-CV1	Oval-CV2
N_1-N_4	11,7,3,3	15,9,7,4	16,8,10,3	14,3,10,3	25,9,25,4	21,8,22,4
$\sum_j N_j / N$	0.23	0.34	0.36	0.29	0.61	0.53
Treatment Effect β_d by OLS and Other Estimators						
OLS	1.385* (0.18,2.56)	0.620+ (-0.14,1.47)	0.655 (-0.26,1.50)	0.945+ (-0.18,2.01)	0.447 (-0.22,1.05)	0.480 (-0.35,1.19)
BW		-0.052 (-0.56,0.81)	-0.014 (-0.54, 1.22)		0.161 (-0.62,0.74)	-0.238 (-0.80,0.68)
MIN	0.080 (-0.34,0.58)	0.059 (-0.39,0.42)	0.006 (-0.36,0.54)			
RD1	0.448 (-0.75,1.87)	0.272 (-0.60,1.22)	0.128 (-0.71,1.62)			
RD2	0.114 (-0.36,0.84)	0.128 (-0.38,0.70)	0.140 (-0.34,0.66)			
Partial Effect β_1 and β_2 by OLS						
β_1	-0.584 (-1.62,0.39)	-0.382 (-1.24,0.31)	-0.571 (-1.36,0.51)	-0.850 (-1.57,0.36)	-0.551 (-1.21,0.17)	-0.631 (-1.28,0.17)
β_2	-0.627+ (-1.56,0.03)	-0.382+ (-1.13,0.09)	-0.462+ (-1.13,0.03)	-0.719* (-1.38,-0.08)	-0.378* (-0.74,-0.01)	-0.375+ (-0.81,0.01)

Sq, square-neighbor kernel; Oval, oval-neighbor kernel
 RT, rule-of-thumb bandwidth; CV1, CV with 1 bandwidth; CV2, CV with 2
 95% bootstrap CI in (-); *, + for 5,10% level significance
 BW, boundary weight; min, $\min(S_1, S_2)$; RD1, $S_1|\delta_2 = 1$; RD 2, $S_2|\delta_1 = 1$

for Sq-RT, Sq-CV, and Oval-RT, because the existing methods are inconsistent if partial effects are present. The partial effect β_1 is insignificant in all cases.

Table 3. MLI Estimates for Treatment and Partial Effects.

	Sq-RT	Sq-CV1	Sq-CV2	Oval-RT	Oval-CV1	Oval-CV2
$N_1 \sim N_4$	11,7,3,3	15,9,7,4	20,12,12,4	14,3,10,3	22,8,20,4	21,8,22,4
$\sum_j N_j / N$	0.23	0.34	0.46	0.29	0.52	0.53
Treatment Effect β_d by OLS and Other Estimators						
OLS	0.629 (-0.75,1.82)	0.497 (-0.42,1.28)	0.361 (-0.13,1.17)	0.666 (-0.60,1.64)	0.268 (-0.65,1.11)	0.264 (-0.76,1.13)
BW		-0.011 (-0.77,0.71)	-0.127 (-0.49,0.36)		-0.108 (-0.59,0.49)	-0.168 (-0.53,0.47)
MIN	0.197 (-0.09,0.68)	0.242 (-0.21,0.55)	0.242 (-0.21,0.55)			
RD1	0.499 (-0.49,1.46)	0.261 (-0.53,1.00)	-0.229 (-0.69,0.42)			
RD2	0.122 (-0.31,0.62)	0.149 (-0.34,0.57)	0.149 (-0.34,0.57)			
Partial Effect β_1 and β_2 by OLS						
β_1	0.351 (-0.69,1.42)	0.222 (-0.67,1.06)	-0.143 (-0.89,0.60)	0.032 (-1.05,1.01)	-0.146 (-0.90,0.75)	-0.088 (-0.90,0.84)
β_2	-0.412 (-1.17,0.17)	-0.366 (-1.06,0.10)	-0.319 ⁺ (-1.06,0.02)	-0.627 ⁺ (-1.25,0.01)	-0.372 ⁺ (-1.01,0.03)	-0.384 ⁺ (-1.00,0.04)

Sq, square-neighbor kernel; Oval, oval-neighbor kernel
 RT, rule-of-thumb bandwidth; CV1, CV with 1 bandwidth; CV2, CV with 2
 95% bootstrap CI in (-); *, + for 5,10% level significance
 BW, boundary weight; min, $\min(S_1, S_2)$; RD1, $S_1|\delta_2 = 1$; RD 2, $S_2|\delta_1 = 1$

The reader may wonder why the partial effects β_1 and β_2 are negative in Table 2: Would being the majority in either house still help passing bills? For this, recall the slope $E(Y^{01}|S) - E(Y^{00}|S)$ of δ_2 in Equation (3), which shows the effect of passing bills relative to “00,” that is, relative to the Democrats being the majority in both houses. Here, “00” is not really a control in the sense that no treatment is done; rather, it is almost the same treatment as “11.” It is hence natural that the slopes of δ_1 and δ_2 are both negative.

The reader may wonder also why β_d is significantly positive: Would the effect of the Republican majority in both houses not be the same as the Democratic majority to result in $\beta_d = 0$? For this, rewrite the slope of D in Equation (3) as

$$\{E(Y^{11}|S) - E(Y^{00}|S)\} - \{E(Y^{10}|S) - E(Y^{00}|S)\} - \{E(Y^{01}|S) - E(Y^{00}|S)\}:$$

the first term $E(Y^{11}|S) - E(Y^{00}|S)$ might be almost zero due to the symmetry of the either party being the majority in both houses, and the last two terms (i.e., the partial effects) are negative so that the slope of D becomes positive.

In Table 3 for MLI, Sq-CV2, Oval-CV1, and Oval-CV2 use nearly 50% of the available data, and consequently we would trust the other columns (Sq-RT, Sq-CV1, and Oval-RT) more, where the treatment effects fall in 0.50–0.67 which are statistically insignificant though, differently from Table 2. These effect numbers differ much from the estimates from the existing methods in the literature. This difference is understandable in view of the partly significant partial effect β_2 ranging over -0.37–0.63 in the columns for Sq-RT, Sq-CV1, and Oval-RT. The partial effect β_1 is insignificant in all cases as in Table 2.

A simple informative “back-of-the-envelope” calculation comes from positing

$$\begin{aligned} & \beta_s\{\delta_1(1 - \delta_2) + (1 - \delta_1)\delta_2\} + \beta_u\{\delta_1\delta_2 + (1 - \delta_1)(1 - \delta_2)\} \\ & = \beta_u + (\beta_s - \beta_u)\delta_1 + (\beta_s - \beta_u)\delta_2 + 2(\beta_u - \beta_s)\delta_1\delta_2 \end{aligned}$$

where β_s is the effect of the split congress, and β_u is the effect of the united congress. Recall that the slope $2(\beta_u - \beta_s)$ of $D = \delta_1 \delta_2$ is $0.62 \sim 1.39$ in Table 2, and the slope $\beta_s - \beta_u$ of δ_2 is $-0.38 \sim -0.72$. Taking the middle values in these ranges, since $2(\beta_u - \beta_s) \approx 1.01$ and $\beta_s - \beta_u \approx -0.55$, we have $\beta_u \approx \beta_s + 0.5$: the effect of the united congress might be about $0.5 \times \text{SD}$ greater than the effect of the split congress.

One final important point to make is that, even if one is interested only in the effect of being the majority in both houses, it is ill-advised to compare only the cases of being the majority in both houses versus not being the majority in neither house. This amounts to omitting δ_1 and δ_2 in the above OLS, which results in an omitted variable bias, as long as the partial effects are not zero as in Tables 2 and 3.

5 Conclusions

In this paper, we generalized the usual mean-based RD with a single running variable (“score”) in three ways by allowing for (i) more than one scores, (ii) partial effects due to part of the scores crossing cutoff, in addition to the full effect with all scores crossing all cutoffs, and (iii) regression functions other than the mean although we focused mostly on the mean. The critical difference between our and existing approaches for MRD is partial effects: allowed in this paper, but ruled out in most other papers.

We imposed a weak continuity assumption, presented the identified parameters, and proposed simple local difference-in-differences-type estimators implemented by ordinary least squares estimator. We applied our estimators to find the US congress “productivity”: the effect of the Republicans dominating both houses on passing bills. We found significant partial effects, and the legislative productivity is higher by about $0.5 \times \text{SD}$ when the congress is united than divided.

Appendix. Three-Score MRD Identification

Consider $D = \delta_1 \delta_2 \delta_3$ with $\delta_j \equiv 1[0 \leq S_j], j = 1, 2, 3$. Rewrite $E(Y|S)$ as

$$\begin{aligned} E(Y|S) &= E(Y^{000}|S)(1 - \delta_1)(1 - \delta_2)(1 - \delta_3) + E(Y^{100}|S)\delta_1(1 - \delta_2)(1 - \delta_3) \\ &+ E(Y^{010}|S)(1 - \delta_1)\delta_2(1 - \delta_3) + E(Y^{001}|S)(1 - \delta_1)(1 - \delta_2)\delta_3 + E(Y^{110}|S)\delta_1\delta_2(1 - \delta_3) \\ &+ E(Y^{101}|S)\delta_1(1 - \delta_2)\delta_3 + E(Y^{011}|S)(1 - \delta_1)\delta_2\delta_3 + E(Y^{111}|S)\delta_1\delta_2\delta_3. \end{aligned}$$

Here, the slope of $D = \delta_1 \delta_2 \delta_3$ is

$$\begin{aligned} E(Y^{111}|S) - E(Y^{110}|S) - \{E(Y^{011}|S) - E(Y^{010}|S)\} \\ - [E(Y^{101}|S) - E(Y^{100}|S) - \{E(Y^{001}|S) - E(Y^{000}|S)\}]. \end{aligned}$$

Adopt the notation analogous to that for two-score MRD.

Take the triple limits on $E(Y|S)$ to get

$$\begin{aligned} E(Y|0^+, 0^+, 0^+) &= E(Y^{111}|0^+, 0^+, 0^+), & E(Y|0^+, 0^+, 0^-) &= E(Y^{110}|0^+, 0^+, 0^-), \\ E(Y|0^-, 0^+, 0^+) &= E(Y^{011}|0^-, 0^+, 0^+), & E(Y|0^-, 0^+, 0^-) &= E(Y^{010}|0^-, 0^+, 0^-), \\ E(Y|0^+, 0^-, 0^+) &= E(Y^{101}|0^+, 0^-, 0^+), & E(Y|0^+, 0^-, 0^-) &= E(Y^{100}|0^+, 0^-, 0^-), \\ E(Y|0^-, 0^-, 0^+) &= E(Y^{001}|0^-, 0^-, 0^+), & E(Y|0^-, 0^-, 0^-) &= E(Y^{000}|0^-, 0^-, 0^-). \end{aligned}$$

These give the limiting version of the slope of D :

$$\begin{aligned} E(Y|0^+, 0^+, 0^+) - E(Y|0^+, 0^+, 0^-) - \{E(Y|0^-, 0^+, 0^+) - E(Y|0^-, 0^+, 0^-)\} \\ - [E(Y|0^+, 0^-, 0^+) - E(Y|0^+, 0^-, 0^-) - \{E(Y|0^-, 0^-, 0^+) - E(Y|0^-, 0^-, 0^-)\}]. \end{aligned}$$

Assume the continuity conditions

$$\begin{aligned} E(Y^{110}|0^+, 0^+, 0^-) &= E(Y^{110}|0^+, 0^+, 0^+), \\ E(Y^{011}|0^-, 0^+, 0^+) &= E(Y^{011}|0^+, 0^+, 0^+), \quad E(Y^{010}|0^-, 0^+, 0^-) = E(Y^{010}|0^+, 0^+, 0^+), \\ E(Y^{101}|0^+, 0^-, 0^+) &= E(Y^{101}|0^+, 0^+, 0^+), \quad E(Y^{100}|0^+, 0^-, 0^-) = E(Y^{100}|0^+, 0^+, 0^+), \\ E(Y^{001}|0^-, 0^-, 0^+) &= E(Y^{001}|0^+, 0^+, 0^+), \quad E(Y^{000}|0^-, 0^-, 0^-) = E(Y^{000}|0^+, 0^+, 0^+). \end{aligned}$$

With these, the slope of D in the preceding display can be written as, not surprisingly, “difference in differences in differences”:

$$\beta_d = E[Y^{111} - Y^{110} - (Y^{011} - Y^{010}) - \{Y^{101} - Y^{100} - (Y^{001} - Y^{000})\}|0^+, 0^+, 0^+]$$

which is the effect on the just treated ($0^+, 0^+, 0^+$). For four scores or more, we get quadruple or higher differences; see Lee (2016).

References

- Angrist, J. D., and V. Lavy. 1999. Using Maimonides’ rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* 114:533–575.
- Battistin, E., A. Brugiavini, E. Rettore, and G. Weber. 2009. The retirement consumption puzzle: evidence from a regression discontinuity approach. *American Economic Review* 99:2209–2226.
- Bayer, P., F. Ferreira, and R. Mcmillan. 2007. A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy* 115:588–638.
- Bernardi, F. 2014. Compensatory advantage as a mechanism of educational inequality: a regression discontinuity based on month of birth. *Sociology of Education* 87:74–88.
- Black, S. E. 1999. Do better school matter? Parental evaluation of elementary education. *Quarterly Journal of Economics* 114:577–599.
- Broockman, D. E. 2009. Do congressional candidates have reverse coattails? Evidence from a regression discontinuity design. *Political Analysis* 17:418–434.
- Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82:2295–2326.
- Cattaneo, M. D., B. Frandsen, and R. Titiunik. 2015. Randomization Inference in the regression discontinuity design: an application to party advantages in the U.S. Senate. *Journal of Causal Inference* 3(1):1–24.
- Cattaneo, M. D., L. Keele, R. Titiunik, and G. Vazquez-Bare. 2016. Interpreting regression discontinuity designs with multiple cutoffs. *Journal of Politics* 78:1229–1248.
- Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare. 2017. Comparing inference approaches in RD designs: a reexamination of the effect of Head Start on child mortality. *Journal of Policy Analysis and Management* 36:643–681.
- Caughey, D., and J. S. Sekhon. 2011. Elections and the regression discontinuity design: lessons from close U.S. House races 1942–2008. *Political Analysis* 19:385–408.
- Choi, J. Y., and M. J. Lee. 2017. Regression discontinuity: review with extensions. *Statistical Papers* 58:1217–1246.
- Choi, J. Y., and M. J. Lee. 2018. Replication data for: Regression discontinuity with multiple running variables allowing partial effects, doi:10.7910/DVN/9YCVJC, Harvard Dataverse, V1, UNF:6:nmh8KvGK8KXOIEJaGnTL4Q==.
- Clark, D., and P. Martorell. 2014. The signaling value of a high school diploma. *Journal of Political Economy* 122:282–318.
- Dell, M. 2010. The persistent effects of Peru’s mining Mita. *Econometrica* 78:1863–1903.
- Eggers, A. C., A. Fowler, J. Hainmueller, A. B. Hall, and J. M. Snyder Jr. 2015. On the validity of the regression discontinuity design for estimating electoral effects: new evidence from over 40,000 close races. *American Journal of Political Science* 59:259–274.
- Gelman, A., and G. Imbens. 2018. Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business and Economic Statistics*, forthcoming.
- Grant, J. T., and N. J. Kelly. 2008. Legislative productivity of the U.S. Congress 1789–2004. *Political Analysis* 16:303–323.
- Imbens, G. W., and K. Kalyanaraman. 2012. Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies* 79:933–959.

- Imbens, G. W., and T. Lemieux. 2008. Regression discontinuity designs: a guide to practice. *Journal of Econometrics* 142:615–635.
- Imbens, G. W., and T. Zajonc. 2009. Regression discontinuity design with vector-argument assignment rules. Unpublished paper.
- Jacob, B. A., and L. Lefgren. 2004. Remedial education and student achievement: a regression discontinuity analysis. *Review of Economics and Statistics* 86:226–244.
- Keele, L. J., and R. Titiunik. 2015. Geographic boundaries as regression discontinuities. *Political Analysis* 23:127–155.
- Keele, L. J., R. Titiunik, and J. 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)* 178:223–239.
- Kemp, G. C. R., and J. M. C. Santos-Silva. 2012. Regression towards the mode. *Journal of Econometrics* 170:92–101.
- Koenker, R. 2005. *Quantile regression*. New York: Cambridge University Press.
- Lalive, R. 2008. How do extended benefits affect unemployment duration? A regression discontinuity approach. *Journal of Econometrics* 142:785–806.
- Lee, D. S., and T. Lemieux. 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* 48:281–355.
- Lee, M. J. 1989. Mode regression. *Journal of Econometrics* 42:337–349.
- Lee, M. J. 2016. *Matching, regression discontinuity, difference in differences and beyond*. Oxford: Oxford University Press.
- Ludwig, J., and D. Miller. 2007. Does head start improve children's life chances? evidence from a regression discontinuity design. *Quarterly Journal of Economics* 122:159–208.
- Matsudaira, J. D. 2008. Mandatory summer school and student achievement. *Journal of Econometrics* 142:829–850.
- Rao, H., L. Q. Yu, and P. Ingram. 2011. Laws of attraction: regulatory arbitrage in the face of activism in right-to-work states. *American Sociological Review* 76:365–385.
- Reardon, S. F., and J. P. Robinson. 2012. Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness* 5:83–104.
- Thistlethwaite, D., and D. T. Campbell. 1960. Regression discontinuity analysis: an alternative to the ex post facto experiment. *Journal of Educational Psychology* 51:309–317.
- Van der Klaauw, W. 2002. Estimating the effect of financial aid offers on college enrollment: a regression discontinuity approach. *International Economic Review* 43:1249–1287.
- Wong, V. C., P. M. Steiner, and T. D. Cook. 2013. Analyzing regression discontinuity designs with multiple assignment variables: a comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics* 38:107–141.