Regression Discontinuity with Multiple Running Variables Allowing Partial Effects

(August 10, 2015)

Jin-young Choi
Faculty of Economics and Business Administration
Goethe University
Frankfurt, Germany
choi@econ.uni-frankfurt.de

Myoung-jae Lee*
Department of Economics
Korea University
Seoul 136-701
South Korea
myoungjae@korea.ac.kr

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' three ways: (i) considering multiple scores, (ii) accommodating quantile/mode regressions, and (iii) allowing 'partial effects' due to each score crossing its own cutoff, not just the full effect with all scores crossing all cutoffs. This generalization is motivated by (i) many multiple-score RD cases, (ii) informative quantile/mode regression functions, and (iii) the full-effect identification needing the partial effects to be separated. We establish identification for '*multiple-score RD (MRD)*', and propose simple estimators that become 'local difference in differences (DD)' in case of double score. We also provide an empirical illustration with German data where partial effects exist for quantile regressions.

Running Head: Multiple-Score RD with Partial Effects.
Key words: regression discontinuity, multiple running variable, quantile, mode, partial effect, difference in differences.

* Corresponding author: 82-2-3290-2229 (phone) and 82-2-926-3601 (fax).

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 new, Rao et al. (2011) and Bernardi (2014) in sociology; Henry et al. (2010) and Lipsey et al. (2015) in education; Chen and Shapiro (2007) and Berk et al. (2010) in criminology; Broockman (2009), 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) and Lee and Lemieux (2010).

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 and 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 et al. 2007, and Michalopoulos and Papaioannou 2014). More examples with multiple running variables will be seen later. 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, and Wong et al. 2013), and 'AND case' where all scores should cross all cutoffs to get treated as in the above spatial RD examples. For simplification, we will examine only AND cases in this paper, because an OR case can be converted to the AND case by "flipping" the treatment, i.e., by relabeling the treatment and control groups.

'*Multiple-score RD for a single treatment (MRD)*' that is the focus of this paper differs from 'RD with multiple cutoffs for a single score' as in Van der Klaauw (2002) and Angrist and Lavy (1999), which is easily handled by looking at each cutoff one at a time. Multiple-score RD for a single treatment also differs from 'multiple-score RD for multiple treatments' as in Leuven et al. (2007), Papay et al. (2011) and Abdulkadiroglu et al. (2014) where each score dictates one treatment, which is a special case of multiple treatments (see, e.g., Lee 2005).

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, as was just mentioned. Second, instead of the usual mean regression function, we will look at a general regression function, say $g(S)$, that includes S-conditional quantiles and mean; $g(S)$ can be even mode (Lee 1989, Kemp and Santos-Silva 2012, and Yao and Li 2014). Allowing quantiles and mode is convenient, because quantile regression (Koenker 2005) can deal with censored responses semiparametrically (Powell 1986) and mode regression can handle truncated responses (Lee 1989) whereas mean cannot. Third, 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.

Certainly, we are not the first to deal with MRD theoretically. Wong et al. (2013) examined 'OR-case MRD', and Keele and Titiunik (2015) 'AND-case MRD'; there are also other papers to be examined later. The critical difference between these studies and this paper is that we allow partial effects while they do not. To see the point, consider two-dimensional score $S = (S_1, S_2)'$ and

$$g(S) = \beta_0 + \beta_1\delta_1 + \beta_2\delta_2 + \beta_d D, \qquad \delta_j \equiv 1[c_j \leq S_j],\ j = 1, 2 \quad \& \quad D = \delta_1\delta_2 \qquad (1.1)$$

where $c \equiv (c_1, c_2)'$ are known cutoffs, and $1[A] = 1$ if A holds and 0 otherwise. Since D is the interaction of δ_1 and δ_2, it is a common practice to allow partial effects β_1 and β_2 of δ_1 and δ_2 in the model. For instance, in the school graduation 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 person. Ruling out such partial effects, Wong et al. (2013) and Keele and Titiunik (2015) found "boundary-specific" effects (which are then to be weighted-averaged), in comparison to our simple effect at $S = c$ (under a weak continuity condition only at $S = c$). We will explain this point in detail later.

Yet another difference between this paper and the papers in the multiple-score RD literature is that we allow quantile/mode effects, whereas almost all those papers considered only mean effects. Generalizing mean effects into quantile/mode effects is not easy, as mean is a linear operator while quantiles/mode are not. This level of generality, however, comes only under 'sharp RD' where D is fully determined by the scores; hence, we stick to sharp MRD in this paper, as Wong et al. (2013) and Keele and Titiunik (2015) also did. For simplification, we will examine only two scores $S = (S_1, S_2)'$, as generalizations to more than two scores are conceptually straightforward. Without loss of generality, we will set the cutoffs at zero unless otherwise noted, as (S_1, S_2) can be always centered as $(S_1 - c_1, S_2 - c_2)$.

In short, *we examine AND-case two-score sharp MRD allowing partial effects for a regression function $g(S)$ that can be the S-conditional mean, quantile or mode. Since the treatment D takes the interaction form $\delta_1\delta_2$ as in (1.1), 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.*

The rest of this paper is organized as follows. Section 2 studies one-score RD for $g(S)$ to set the stage. Section 3 examines the identification and estimation for two-score MRD for $g(S)$—this is the main section of this paper. Section 4 compares our identification conditions and estimators with those in the literature. Section 5 provides an empirical illustration for the effects of unemployment insurance benefit (UIB) on unemployment duration using German data. Finally, Section 6 concludes.

2 RD with One Score for General Regression

Suppose we have a sharp treatment D that equals

$$\delta \equiv 1[0 \leq S].$$

Denoting a value that S can take as s and writing $g(Y|S=s)$ just as $g(Y|s)$, define

$$g(Y|0^+) \equiv \lim_{s\downarrow 0} g(Y|s) \qquad \text{and} \qquad g(Y|0^-) \equiv \lim_{s\uparrow 0} g(Y|s).$$

Let (Y^0, Y^1) denote the potential responses corresponding to $D = \delta = \{0,1\}$, and let the observed response be $Y = (1-D)Y^0 + DY$. Rewrite $g(Y|S)$ as

$$\begin{aligned} g(Y|S) &= g(Y^0|S)(1-D) + g(Y^1|S)D = g(Y^0|S) + \{g(Y^1|S) - g(Y^0|S)\}D \\ &= g(Y^0|S) + \beta(S)D \qquad \text{where} \qquad \beta(S) \equiv g(Y^1|S) - g(Y^0|S). \end{aligned} \tag{2.1}$$

Assume that $g(Y^0|s)$ *is continuous at* 0 *and* $g(Y^1|0^+)$ *exists*; the latter is hardly an assumption. Taking the upper/right and lower/left limits at 0 on (2.1) gives

$$g(Y|0^+) = g(Y^0|0^+) + \beta(0^+) \qquad \text{and} \qquad g(Y|0^-) = g(Y^0|0^-).$$

Subtract the latter from the former, and invoke $g(Y^0|0^+) = g(Y^0|0^-)$ to obtain

$$\begin{aligned} \beta_d &\equiv \beta(0^+) = g(Y|0^+) - g(Y|0^-) \qquad (2.2) \\ \{ &= g(Y^1|0^+) - g(Y^0|0^-) = g(Y^1|0^+) - g(Y^0|0^+) \ \text{ as } g(Y^0|s) \text{ is continuous at } 0\}. \end{aligned}$$

The (local at $s=0$) treatment effect β_d is identified by $g(Y|0^+) - g(Y|0^-)$, and β_d is characterized by $g(Y^1|0^+) - g(Y^0|0^+)$ that becomes $E(Y^1 - Y^0|0^+)$ when $g(\cdot|S) = E(\cdot|S)$—the mean effect on the 'just treated'.

Differently from the usual mean RD, if $g(\cdot)$ is a quantile function, then

$$\beta_d = g(Y^1|0^+) - g(Y^0|0^+) \neq g(Y^1 - Y^0|0^+) \qquad \text{in general;} \tag{2.3}$$

$g(Y^1 - Y^0|0^+)$ that is the more appropriate quantile effect than $g(Y^1|0^+) - g(Y^0|0^+)$ was examined in Lee (2000), but rarely so in the RD literature. In this paper, we will also adopt the quantile effect in the form $g(Y^1|0^+) - g(Y^0|0^+)$, rather than the more difficult $g(Y^1 - Y^0|0^+)$. The problem (2.3) also occurs for mode.

Equation (2.2) shows how to estimate β_d: obtain sample versions $\hat{g}(Y|S=h)$ and $\hat{g}(Y|S=-h)$ for a small bandwidth $h > 0$, and then use

$$\hat{\beta}_d \equiv \hat{g}(Y|S=h) - \hat{g}(Y|S=-h). \qquad (2,4)$$

Instead of this, however, a much simpler implementation is replacing $\beta(S)$ in (2.1) with β_d and $g(Y^0|S)$ with a linear spline of S to get

$$g(Y|S) = \beta_0 + \beta_- S(1-\delta) + \beta_+ S\delta + \beta_d D \tag{2.5}$$

where β's are parameters, and then applying quantile/mode regression to this.

Equation (2.5) follows the usual mean-RD implementation: although $g(Y^0|S)$ is a constant at $S = 0$, within the local neighborhood $\pm h$ of 0 on which estimation is done in practice, it is better to allow $g(Y^0|S)$ to change at least "linear-splinely"; see Hahn et al. (2001). Instead of a linear spline, a polynomial function may be used such as a quadratic or cubic function of S. How to choose h is still a vexing problem in the usual mean RD, despite theoretical advances as in Imbens and Kalyanaraman (2012) and Calonico et al. (2014).

Finally in this section for a scalar S, we note that Frandsen et al. (2012) examined quantile effects for fuzzy RD and proposed 'complier quantile effects'. It is not clear, however, how their approach could be generalized to MRD. Instead of this, as can be seen in the next section, our approach generalizes (2.1) that holds only for sharp RD.

3 MRD with Two Scores

Turning to two-score MRD, let

$$S = (S_1, S_2)' \quad \text{and} \quad D = \delta_1\delta_2 \quad \text{where} \quad \delta_j \equiv 1[0 \le S_j],\ j = 1, 2.$$

First, we introduce four potential responses corresponding to $\delta_1 = 0, 1$ and $\delta_2 = 0, 1$, and examine partial effects—an issue that does not arise for a scalar S. Second, we impose a continuity condition analogous to the continuity of $g(Y^0|S)$ at 0 in (2.1), and present the identified effect analogous to $g(Y^1|0^+) - g(Y^0|0^+)$ in (2.2). Third, we propose estimators that differ in how a baseline function analogous to $g(Y^0|S)$ in (2.1) is specified.

3.1 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 $Y^{11} - Y^{00}$.

Recalling (2.1), rewrite $g(Y|S)$ as

$$g(Y|S) = g(Y^{00}|S)(1-\delta_1)(1-\delta_2) + g(Y^{10}|S)\delta_1(1-\delta_2) + g(Y^{01}|S)(1-\delta_1)\delta_2 + g(Y^{11}|S)\delta_1\delta_2. \tag{3.1}$$

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

$$\begin{aligned} g(Y|S) \;=\; & g(Y^{00}|S) + \{g(Y^{10}|S) - g(Y^{00}|S)\}\delta_1 + \{g(Y^{01}|S) - g(Y^{00}|S)\}\delta_2 \\ & + \{g(Y^{11}|S) - g(Y^{10}|S) - g(Y^{01}|S) + g(Y^{00}|S)\}D \end{aligned} \tag{3.2}$$

which will play the main role for MRD. This equation does not hold for fuzzy RD as is the case for (2.1), because D would then depend on random variables other than S

on the right-hand side while the left-hand side $g(Y|S)$ is a function of only S. This is why we stick to sharp RD.

The slope of $D = \delta_1\delta_2$ in (3.2) is reminiscent of the above $Y^{11} - Y^{10} - Y^{01} + Y^{00}$, and it is a DD with $g(Y^{11}|S) - g(Y^{10}|S)$ as the "treatment group difference" and $g(Y^{01}|S) - g(Y^{00}|S)$ 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

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

then (3.2) becomes

$$g(Y|S) = g(Y^{00}|S) + \beta^{00}(S)D \quad \text{where} \quad \beta^{00}(S) \equiv g(Y^{11}|S) - g(Y^{00}|S) \tag{3.3}$$

that is analogous to the single-score equation (2.1); '00' in $\beta^{00}(S)$ refers to the baseline superscript in Y^{00}.

It helps to see when the no partial-effect assumption is violated: recall (1.1) with $\beta_1 \neq 0$ or $\beta_2 \neq 0$, where

$$\begin{aligned} & g(Y^{11}|s) = \beta_0 + \beta_1 + \beta_2 + \beta_d, \qquad g(Y^{10}|s) = \beta_0 + \beta_1, \\ & g(Y^{01}|s) = \beta_0 + \beta_2, \qquad\qquad\qquad g(Y^{00}|s) = \beta_0 \\ \Longrightarrow\ & g(Y^{11}|s) - g(Y^{10}|s) - g(Y^{01}|s) + g(Y^{00}|s) = \beta_d, \\ & g(Y^{11}|s) - g(Y^{00}|s) = \beta_1 + \beta_2 + \beta_d. \end{aligned} \tag{3.4}$$

Examine squares 1∼4 in Figure 1, where (h_1, h_2) are the 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, squares 1 and 4, or squares 1 and 3. With partial effects present, however, this is no longer the case: squares 1 and 2 give the treatment effect plus the partial effect due to S_1 crossing 0; squares 1 and 4 give the treatment effect plus the partial effect due to S_2 crossing 0; squares 1 and 3 gives the treatment effect plus the two partial effects. It is only when we take DD as in (3.2) that the desired treatment effect is identified.

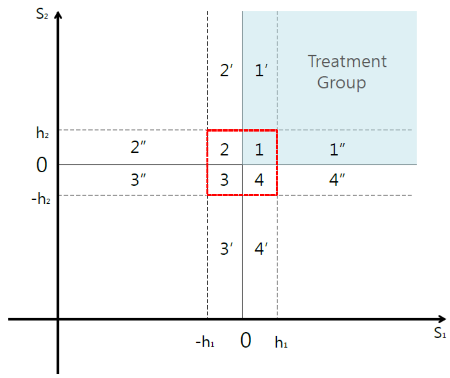

Figure 1: Two Score RD in 'AND' Case

3.2 Identification

To simplify notation, denote

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

With little loss of generality, assume that these double limits of $g(\cdot|S)$ exist at 0 for the potential responses, and denote them using 0^- and 0^+; e.g.

$$g(Y^{00}|0^-, 0^+) \equiv \lim_{-,+} g(Y^{00}|s_1, s_2).$$

Take the double limits on (3.1) to get

$$\begin{aligned} g(Y|0^+, 0^+) &= g(Y^{11}|0^+, 0^+), \qquad g(Y|0^+, 0^-) = g(Y^{10}|0^+, 0^-), \\ g(Y|0^-, 0^+) &= g(Y^{01}|0^-, 0^+), \qquad g(Y|0^-, 0^-) = g(Y^{00}|0^-, 0^-). \end{aligned}$$

These give a limiting version of the slope of $D = \delta_1 \delta_2$ in (3.2) at $(0, 0)$:

$$\begin{aligned} & g(Y|0^+, 0^+) - g(Y|0^+, 0^-) - g(Y|0^-, 0^+) + g(Y|0^-, 0^-) \\ = \; & g(Y^{11}|0^+, 0^+) - g(Y^{10}|0^+, 0^-) - g(Y^{01}|0^-, 0^+) + g(Y^{00}|0^-, 0^-). \qquad (3.5) \end{aligned}$$

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

$$\begin{aligned}(i) &: g(Y^{01}|0^-, 0^+) = g(Y^{01}|0^+, 0^+), \\ (ii) &: g(Y^{10}|0^+, 0^-) = g(Y^{10}|0^+, 0^+), \\ (iii) &: g(Y^{00}|0^-, 0^-) = g(Y^{00}|0^+, 0^+).\end{aligned} \tag{3.6}$$

In (3.6), *(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.* Under (3.6), (3.5) becomes

$$g(Y|0^+, 0^+) - g(Y|0^+, 0^-) - g(Y|0^-, 0^+) + g(Y|0^-, 0^-) \tag{3.7}$$

$$= \beta_d \equiv g(Y^{11}|0^+, 0^+) - g(Y^{10}|0^+, 0^+) - g(Y^{01}|0^+, 0^+) + g(Y^{00}|0^+, 0^+) \tag{3.8}$$

that is the correct two-score MRD analog of $g(Y^1|0^+) - g(Y^0|0^+)$ in (2.2).

The expression (3.7) is the identified entity that is characterized by (3.8). When $g(\cdot|S) = E(\cdot|S)$, (3.8) is the effect on the just treated $E(Y^{11} - Y^{10} - Y^{01} + Y^{00}|0^+, 0^+)$. Since each limit exists in (3.8), (3.8) can be written also as

$$\beta_d = \lim_{+,+}\{g(Y^{11}|s) - g(Y^{10}|s) - g(Y^{01}|s) + g(Y^{00}|s)\}.$$

We summarize this main finding (as well as (3.3) under no partial effect) as a theorem.

THEOREM 1: Suppose the double limits of $g(\cdot|S)$ exist at 0 for the potential responses, the continuity condition (3.6) holds, and the density function $f_S(s)$ of S is strictly positive on an neighborhood of $(0, 0)$. Then the effect

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

is identified by two-score MRD (3.7). If no partial-effect condition holds locally at $S = 0$ in the sense $g(Y^{10}|0^+, 0^+) = g(Y^{01}|0^+, 0^+) = g(Y^{00}|0^+, 0^+)$, then β_d becomes $g(Y^{11}|0^+, 0^+) - g(Y^{00}|0^+, 0^+)$.

3.3 Estimation

Define

$$\delta_j^- \equiv 1[-h_j < S_j < 0], \qquad \delta_j^+ \equiv 1[0 \le S_j < h_j], \qquad j = 1, 2.$$

Although (3.7) shows that β_d can be estimated by replacing the four identified elements in (3.7) with their sample versions, in practice, it is easier to implement MRD with (3.2), using only the local observations satisfying $S_1 \in (-h_1, h_1)$ and $S_2 \in (-h_2, h_2)$.

Specifically, replace $g(Y^{00}|S)$ in (3.2) 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

$$g(Y|S) = g(Y^{00}|S) \ + \ \beta_1\delta_1 + \beta_2\delta_2 + \beta_d D \tag{3.9}$$

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

$$\begin{aligned} linear \ &: \ m_1(S) \equiv \text{linear function of } S_1, S_2, &(3.10)\\ quadratic \ &: \ m_2(S) \equiv m_1(S) + \text{linear function of } S_1^2, S_2^2, S_1S_2, \\ cubic \ &: \ m_3(S) \equiv m_2(S) + \text{linear function of } S_1^3, S_2^3, S_1^2S_2, S_1S_2^2. \end{aligned}$$

Then quantile/mode regression can be applied to (3.9). For the mean effect, least squares estimator (LSE) is enough.

Yet another way to specify $g(Y^{00}|S)$ in (3.9) is a piecewise-linear function continuous at 0 as in (2.5):

$$\begin{aligned} g(Y^{00}|S) = \beta_0 &+ \beta_{11}\delta_1^-\delta_2^-S_1 + \beta_{12}\delta_1^-\delta_2^-S_2 + \beta_{21}\delta_1^-\delta_2^+S_1 + \beta_{22}\delta_1^-\delta_2^+S_2 \\ &+\beta_{31}\delta_1^+\delta_2^-S_1 + \beta_{32}\delta_1^+\delta_2^-S_2 + \beta_{41}\delta_1^+\delta_2^+S_1 + \beta_{42}\delta_1^+\delta_2^+S_2. \end{aligned} \tag{3.11}$$

This function linear in S allows different slopes across the four quadrants determined by $(\delta_1^-, \delta_1^+, \delta_2^-, \delta_2^+)$; it is continuous at 0 because $\lim_{s\to 0} g(Y^{00}|s) = \beta_0$ for any sequence of s approaching 0.

4 Other Approaches in the Literature

Having presented our proposals, now we turn to the other approaches in the MRD literature. First, two scores are collapsed into a single score so that the familiar single-score RD arsenal can be mobilized. Second, two-dimensional localization is avoided by doing, e.g., one-dimensional localization for S_1 given $S_2 \geq 0$ (i.e., given $\delta_2 = 1$) to get the 'effects on the boundary $S_2 \geq 0$'; here as well, the familiar single-score RD methods can be utilized. Third, those effects on the boundary can be combined using weights as in Wong et al. (2013) and Keele and Titiunik (2015).

4.1 Minimum of Normalized Scores

Battistin et al. (2009) and Clark and Martorell (2014) defined, for some scale constants σ_1 and σ_2,

$$S_m \equiv \min(\frac{S_1}{\sigma_1}, \frac{S_2}{\sigma_2}) \implies D = 1[0 \leq S_m]$$

to set up

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

where β_m is the treatment effect of interest. Recalling (3.9) with $\beta_1 = \beta_2 = 0$, we can see that $g(Y^{00}|S_1, S_2)$ in (3.9) is specified just as $\beta_0 + \beta_- S_m(1-D) + \beta_+ S_m D$.

This approach has a couple of problems. First, S_m looks like reducing the localization dimension to one, but that is not the case, as there are two actual bandwidths $h_1 = \sigma_1 h_m$ and $h_2 = \sigma_2 h_m$ where h_m is a bandwidth for S_m. Second, the linear spline $\beta_0 + \beta_- S_m(1-D) + \beta_+ S_m D$ is inadequate, as it approximates $g(Y^{00}|S)$ only with S_1 when $S_1/\sigma_1 < S_2/\sigma_2$ and only with S_2 when $S_2/\sigma_2 < S_1/\sigma_2$: there is no reason to voluntarily "handcuff" oneself this way, and better approximations can be seen in (3.10) and (3.11). Third, the most important is that, as in the other approaches, 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 (or S_2 only) is allowed.

A couple of remarks is in order. First, Wong et al. (2013) called this way of dealing with MRD 'centering approach', but 'normalizing approach' (in the sense of location and scale normalization) or 'min approach' would be more fitting. Second, Battistin et al. (2009) and Clark and Martorell (2014) dealt with fuzzy mean-based MRD's, not sharp MRD. Third, using the minimum of the two normalized scores can be easily generalized to more than two scores, as in Clark and Martorell (2014) who had three scores (e.g., three tests to pass) and thus $S_m = \min(S_1/\sigma_1, S_2/\sigma_2, S_3/\sigma_3)$.

4.2 One-Dimensional Localization

The dominant approach in the MRD literature has been looking at a subpopulation with one score already greater than its cutoff (Jacob and Lefgren 2004, Lalive 2008, Matsudaira 2008, Schmieder et al. 2012, and Caliendo et al. 2013). For instance, on the subpopulation with $\delta_1 = 1$, δ_2 equals D, and squares 1 and 1" in Figure 1 become the treatment group whereas squares 4 and 4" become the control group. This will raise estimation efficiency as only one-dimensional localization is done with the larger control and treatment groups, but a bias will appear if there is a partial effect.

To formalize the idea, set $\delta_1 = 1$ ($\Longleftrightarrow S_1 \geq 0$) in (3.2) to have

$$g(Y|S) = g(Y^{10}|S) + \{g(Y^{11}|S) - g(Y^{10}|S)\}\delta_2; \qquad (4.1)$$

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

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

Assume the continuity condition

$$g(Y^{10}|s_1, 0^+) = g(Y^{10}|s_1, 0^-) \qquad \forall s_1 \geq 0; \qquad (4.2)$$

whereas this has '$\forall s_1 \geq 0$', (ii) of (3.6) is only for $s_1 = 0^+$ that is weaker than (4.2). Using (4.2), the difference between the upper and lower limits is

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

'10' in $\beta^{10}(s_1, 0^+)$ refers to the baseline superscript in Y^{10}. For (1.1), $\beta^{10}(s_1, 0^+) = \beta_2 + \beta_d$, not β_d.

Proceeding analogously, set $\delta_2 = 1$ ($\Longleftrightarrow S_2 \geq 0$) in (3.2) to have

$$g(Y|S) = g(Y^{01}|S) + \{g(Y^{11}|S) - g(Y^{01}|S)\}\delta_1; \tag{4.3}$$

$g(Y^{01}|S)$ is the baseline. Take the upper and lower limits only for s_1, with $s_2 \geq 0$:

$$\begin{aligned} g(Y|0^+, s_2) &= g(Y^{01}|0^+, s_2) + \lim_{s_1 \downarrow 0}\{g(Y^{11}|s_1, s_2) - g(Y^{01}|s_1, s_2)\}, \\ g(Y|0^-, s_2) &= g(Y^{01}|0^-, s_2). \end{aligned}$$

Assume the continuity condition

$$g(Y^{01}|0^+, s_2) = g(Y^{01}|0^-, s_2) \qquad \forall s_2 \geq 0; \tag{4.4}$$

whereas this has '$\forall s_2 \geq 0$', (i) of (3.6) is only for $s_2 = 0^+$. Using (4.4), the difference between the upper and lower limits is

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

For (1.1), $\beta^{01}(0^+, s_2) = \beta_1 + \beta_d$, not β_d.

Estimation with one-dimensional localization is simple because there is only one score. Analogously to (3.9), consider for (4.1):

$$g(Y|S) = g(Y^{10}|S) + \beta^{10}\delta_2 \tag{4.5}$$

where $g(Y^{10}|S)$ can be specified as in (3.10) or more generally as in

$$g(Y^{10}|S) = \beta_0 + \beta_{2-}S_2(1 - \delta_2) + \beta_{2+}S_2\delta_2 + \beta_1 S_1;$$

no slope difference for S_1 as $S_1 \geq 0$. Quantile or mode regression can be applied to this using only the subsample with $(\delta_2^- + \delta_2^+)\delta_1 = 1$ (i.e., only S_2 is localized given $S_1 \geq 0$).

We omit the opposite case of smoothing with S_1 given $S_2 \geq 0$; the analog for (4.5) is $g(Y|S) = g(Y^{01}|S) + \beta^{01}\delta_1$, and only the subsample with $(\delta_1^- + \delta_1^+)\delta_2 = 1$ is to be used for estimation. Other than for the bias due to partial effects, estimators for this and (4.5) with one-dimensional localization should be more efficient than those for (3.9) with two-dimensional localization.

4.3 Weighted Average of Boundary Effects

Although many papers dealt with MRD, as far as we are aware of, only Wong et al. (2013) and Keele and Titiunik (2015) examined identification conditions and estimation theoretically. Here we examine the two papers one by one. Since Keele and Titiunik (2015; "KT" in this section) addressed AND-case two-score sharp MRD whereas Wong et al. (2013) did OR-case two-score sharp MRD, we discuss the former first. We will denote the two cutoffs as (c_1, c_2) in this subsection.

Consider the two boundary lines B stemming from (c_1, c_2) rightward and upward in Figure 1. KT assumed away partial effects so that 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 (p. 131) assumed the continuity at all points in B for the potential untreated and treated responses Y_0 and Y_1:

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

where $Y_0 = (Y^{00}, Y^{01}, Y^{10})$ and $Y_1 = Y^{11}$ in our notation.

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 $b \in B$:

$$\begin{aligned}
&\lim_{s^t \to b} E(Y|S = s^t) - \lim_{s^c \to b} E(Y|S = s^c) = \lim_{s^t \to b} E(Y_1|S = s^t) - \lim_{s^c \to b} E(Y_0|S = s^c) \\
= \;& E(Y_1|S = b) - E(Y_0|S = b) = E(Y_0 - Y_0|S = b) \equiv \tau(b). \qquad (4.6)
\end{aligned}$$

A marginal effect can be found by integrating out b: with $f_{S|B}$ and f_S denoting the density of $S|B$ and S, respectively,

$$\tau \equiv \int_B \tau(s) f_{S|B}(s) \partial s = \frac{\int_B \tau(s) \cdot f_S(s) \partial s}{\int_B f_S(s) \partial s}.$$

KT (2015) 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 (4.6). The difference of the intercept estimators is then an estimator for $\tau(b)$.

Although ruling out partial effects seems to be minor, this resulted in major differences. First, whereas we looked at the effect at the single point (c_1, c_2), KT looked as the effect on the entire boundary, as the same treatment is administered regardless of the "source" position when the boundary is crossed. Second, whereas our estimator is a simple generalization of the single-score RD estimator as in Hahn et al. (2001), the estimators of KT for $\tau(b)$ and τ are involved. Third, whereas we need two-dimensional smoothing, KT need one-dimensional smoothing as was the case for (4.1) and (4.3); this will enable more efficient estimation if indeed there is no partial effect.

Turning to Wong et al. (2013; "WSC" from Wong, Steiner and Cook in this section), WSC dealt with OR-case two-score sharp MRD where $D = 1$ if $S_1 < c_1$ or $S_2 < c_1$. In Figure 1, the treatment quadrant in KT is now the control quadrant of WSC, and the control quadrants in KT are the treatment quadrants of WSC. As KT did, WSC also ruled out partial effects: "we partition the treatment space into three subspaces, we assume that all cases receive exactly the same treatment (otherwise, more than one potential treatment outcome needs to be considered)" in WSC (p. 111).

The continuity assumption of WSC is analogous to (4.2) and (4.4): for Y_0,

$$\begin{aligned} \lim_{s_1 \downarrow c_1} E(Y_0|S_1 &= s_1, S_2 > c_2) = \lim_{s_1 \uparrow c_1} E(Y_0|S_1 = s_1, S_2 > c_2), \\ \lim_{s_2 \downarrow c_2} E(Y_0|S_1 &> c_1, S_2 = s_2) = \lim_{s_2 \uparrow c_2} E(Y_0|S_1 > c_1, S_2 = s_2); \end{aligned}$$

the same continuity should hold also for Y_1.

WSC laid out four approaches, among which we explain three (the last one does not seem tenable, and WSC did not recommend it either). The first is the minimum of normalized scores, which was examined already. The second is essentially the one-dimensional localization approach along the horizontal boundary (say B_1) of B, and then along the vertical boundary (say B_2)—WSC called this 'univariate approach'; 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—WSC called this 'frontier approach' to dub the weighted average 'frontier average treatment effect'.

Although disallowing partial effects may look simplifying, to the contrary, it resulted in considering the boundary lines B_1 and B_2 instead of the single boundary point (c_1, c_2). The possibly heterogeneous effects along the boundary lines may be informative and possibly efficiency-enhancing, but it also raises the question of finding a single marginal effect as an 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 non-zero partial effects first with our approach, and then test for zero partial effects; if accepted, one may adopt the approaches in KT and WSC. This should be preferred than simply ruling out partial effects from the beginning, unless there is a strong prior justification to do so.

5 Empirical Example: Unemployment Duration

This section provides an empirical illustration for effects of unemployment insurance benefit (UIB) on unemployment duration using German data. Our data set was drawn from the Sample of Integrated Labour Market Biographies 1975-2010 that is a 2% sample from the Integrated Employment Biographies (IEB) of the Institute for Employment Research in Germany. IEB contains daily information on individual employment and UIB. IEB was used also by Schmieder et al. (2012) and Caliendo et al. (2013).

For simplicity, we restrict our sample to the inflow individuals into unemployment for years 2001-2003. The individual employment outcomes are observed up to 3 years after entering unemployment, which resulted in the right-censoring problem at 3 years. Our Y is ln(unemployment duration in months).

UIB duration is determined by *age* S_1 *and the number of working months* S_2 in the last seven years as follows—age is observed only in integer years, unfortunately:

1. For $S_1 < 45$, 6 months if $12 \leq S_2 < 16$, 8 months if $16 \leq S_2 < 20$, 10 months

if $20 \leq S_2 < 24$, and 12 months if $24 \leq S_2$; see the second and third quadrants in Figure 2, where the same color above the $S_2 = 24$ line means the same UIB duration regardless of S_2.

2. For $45 \leq S_1 < 47$, 6 months if $12 \leq S_2 < 16$, 8 months if $16 \leq S_2 < 20$, 10 months if $20 \leq S_2 < 24$, 12 months if $24 \leq S_2 < 28$, 14 months if $28 \leq S_2 < 32$, 16 months if $32 \leq S_2 < 36$, and 18 months if $36 \leq S_2$; see also Figure 2.

3. For $47 \leq S_1 < 52$, there are multiple cutoffs for S_2 at 16, 20, 24, 28, 32, 36, 40, and 44 as can be in Figure 2; there are also multiple cutoffs for S_1 beyond $S_1 \geq 52$, although not visible in Figure 2.

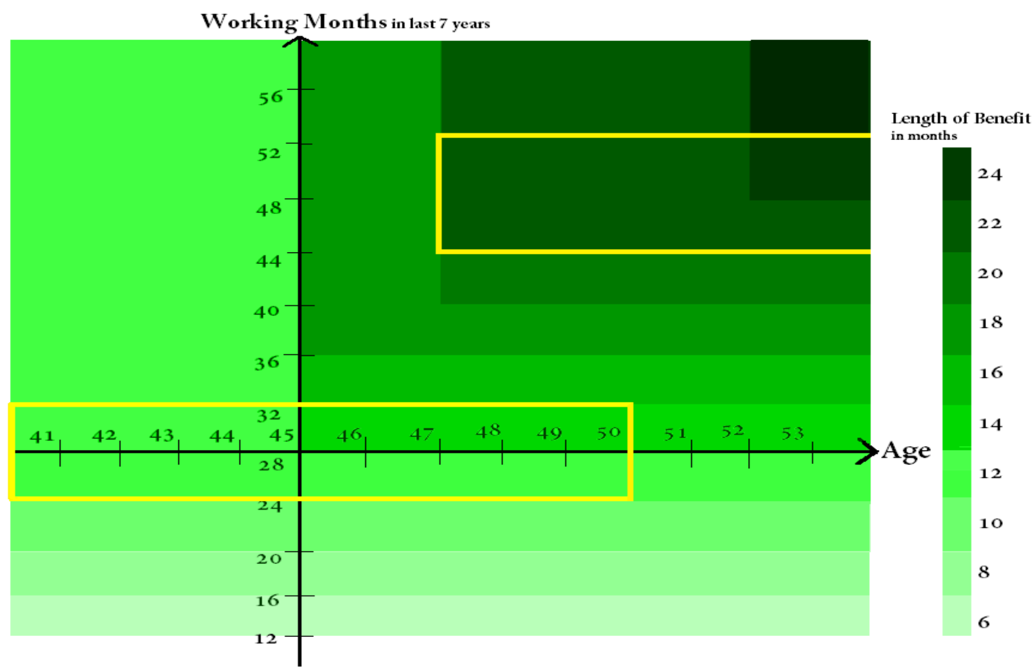

Figure 2: UIB Duration by Age and Working Months

In our empirical analysis, as there are too many treatments depending on different cutoffs, we focus on two cutoffs. The first cutoff $(c_1, c_2) = (45, 28)$ is the "center" in Figure 2, and we use the subsample with ages 45 ± 5 and working months 28 ± 4

(the lower-left yellow box). In the box, the darker quadrant is the treatment group with 14 months of UIB corresponding to square 1 in Figure 1, and the three lighter quadrants are the control group with 12 months of UIB corresponding to squares 2-4. Hence, the treatment is receiving additional 2 months of UIB. The second cutoff is $(c_1, c_2) = (52, 48)$, and we use the subsample with ages 52 ± 5 and working months 48 ± 4 (the upper-right yellow box in Figure 2). In the box, the darker quadrant is the treatment group with 24 months of UIB, and the three lighter quadrants are the control group with 22 months of UIB. Here again, the treatment is receiving additional 2 months of UIB.

Table 1: Data Description

$(45 \pm 5, 28 \pm 4)$, $N = 2145$						
	Total			Non-censored (78%)		
Variable	Mean	Med	SD	Mean	Med	SD
Unemp. dur. in months $\exp(Y)$	17.3	13	13.1	11.9	9	9.45
Age in years (S_1)	43.9	44	2.80	43.9	44	2.80
Working months (S_2)	27.1	27	2.50	27.2	27	2.50
Female	0.47			0.50		
$(52 \pm 5, 48 \pm 4)$, $N = 619$						
	Total			Non-censored (86%)		
Variable	Mean	Med	SD	Mean	Med	SD
Unemp. dur. in months $\exp(Y)$	13.7	9	12.2	10.0	7	8.85
Age in years (S_1)	50.8	51	2.75	50.9	51	2.79
Working months (S_2)	47.2	47	2.26	47.2	47	2.23
Female	0.38			0.35		

Table 1 compares the whole sample with the non-censored subsample for $(c_1, c_2) = (45, 28)$ and $(52, 48)$. The local sample size $N = 2145$ for $(c_1, c_2) = (45, 28)$ is much greater than $N = 619$ for $(c_1, c_2) = (52, 48)$. In the first panel for $(45 \pm 5, 28 \pm 4)$, the average age and working months are the same across the two groups, and the proportion of females is only 3% higher in the non-censored group. Hence, although the recorded unemployment duration is shorter for the obvious reason for the non-censored group, there is little systematic difference between the whole sample and the non-censored subsample locally around $(c_1, c_2) = (45, 28)$. In the second panel for $(52 \pm 5, 48 \pm 4)$, the mean and median unemployment durations are shorter than those in the first panel by $2 \sim 4$ months, and the proportion of the non-censored observations is higher by 8%. As in the first panel, there is little systematic difference between the whole sample and the non-censored subsample, except that the proportion of females is 3% lower in the non-censored group.

Table 2: Unemployment Duration with $(45 \pm 5, 28 \pm 4)$

	Males				Females			
Quantile %	Q1 (Treated)	Q2	Q3	Q4	Q1 (Treated)	Q2	Q3	Q4
5%	1	1	1	1	1	1	1	1
25%	3	3	3	4	6	3	4	5
50%	8	7	9	11	13	7	10	11
75%	14	17	18	20	20	20	18	19
95%	30	30	31	30	29	31	31	30

Table 3: Unemployment Durations with $(52 \pm 5, 48 \pm 4)$

	Males				Females			
Quantile %	Q1 (Treated)	Q2	Q3	Q4	Q1 (Treated)	Q2	Q3	Q4
5%	1	1	1	1	1	1	1	1
25%	4	3	3	2	6	5	3	2
50%	9	7	5	4	9	10	8	4
75%	15	16	14	15	21	17	18	9
95%	28	32	28	26	26	29	30	26

Tables 2 and 3 present quantiles of unemployment durations ($\exp(Y)$) for males and females using only the non-censored observations, where Q1-Q4 denote the four squares in Figure 1. Since only the non-censored observations are used, Tables 2 and 3 are not exactly informative for the population. Nevertheless, Table 2 suggests that the quantile effects at $(45, 28)$ may be different: near zero effects for extreme quantiles, and some effects in "middle" quantiles. Table 3 suggests that the three control groups are heterogeneous to result in partial effects at $(52, 48)$. For instance, for females, Q3 and Q4 differ by 4 at the median and by 9 at the upper quartile. The magnitude 9 is not small, being about one standard deviation (SD) as can be seen in Table 1.

Turning to quantile regression, due to the right-censoring problem, we applied the censored least absolute deviation estimator (CLAD) of Powell (1986) to (3.9) with $q(Y^{00}|S) = m_2(S)$ in (3.10). CLAD does not impose parametric assumptions on the model error term and allows heteroskedasticity of an unknown form. Figure 3 is for $(c_1, c_2) = (45, 28)$, and Figure 4 is for $(52, 48)$. In the figures, the first row is for β_d, the second row for β_1, and the third for β_2. Also, the first column is for the full sample (i.e. the local sample with both males and females), the second for the males only, and the third for the females only. The solid lines are the estimates, and the dashed lines are 90% point-wise confidence intervals calculated by a bias-corrected nonparametric bootstrap with 300 bootstrap pseudo estimates.

Examining Figure 3, for the full sample, the treatment effects are about $0.35 \sim 0.69$ for most quantiles. Since the median duration is about $9 \sim 13$ in Table 1, this translates into about $5 \sim 7$ month increase in unemployment duration. For males, the effects are positive but smaller than those of the full sample and most estimates are insignificant. For females, the effects are significantly positive $(0.69 \sim 0.84)$ over $\alpha = 0.2 \sim 0.6$. Thus the significant effects in the full sample are driven by the females.

Regarding the partial effects, the second row shows that the partial effect β_1 of age is insignificant and close to zero for males. For females, however, some negative effects $(-0.35 \sim -0.4)$ are seen around 0.6 quantile that are nearly significant, which results in smaller negative effects for the full sample. As for the partial effect β_2 of working months, all quantile effects are near zero and insignificant.

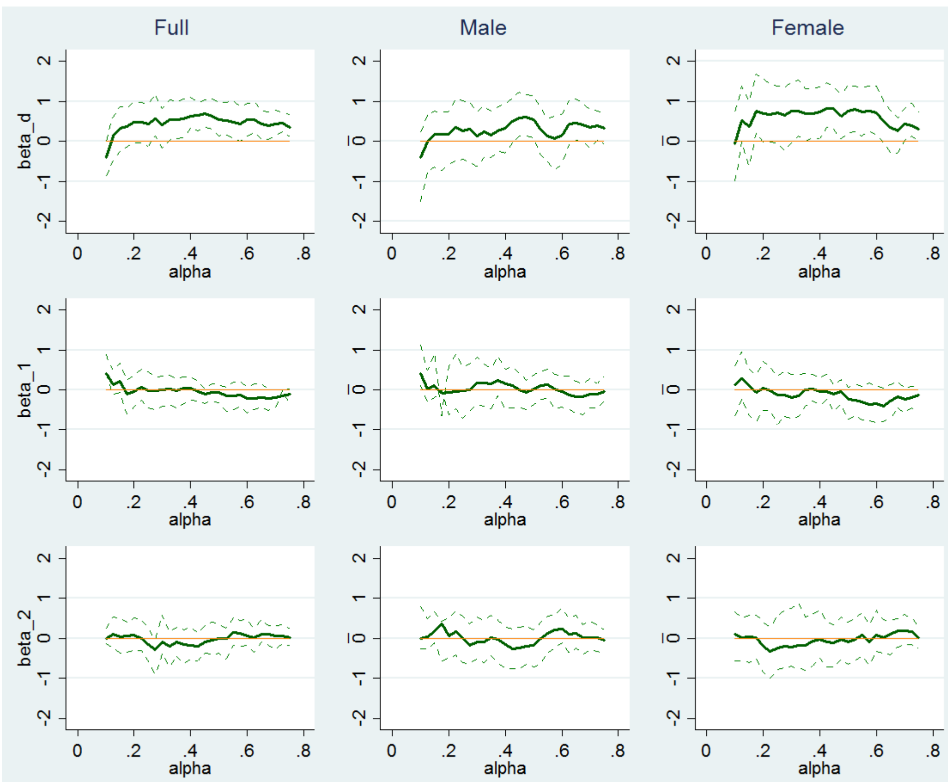

Figure 3: Quantile Effects of UIB at $(c_1, c_2) = (45, 28)$

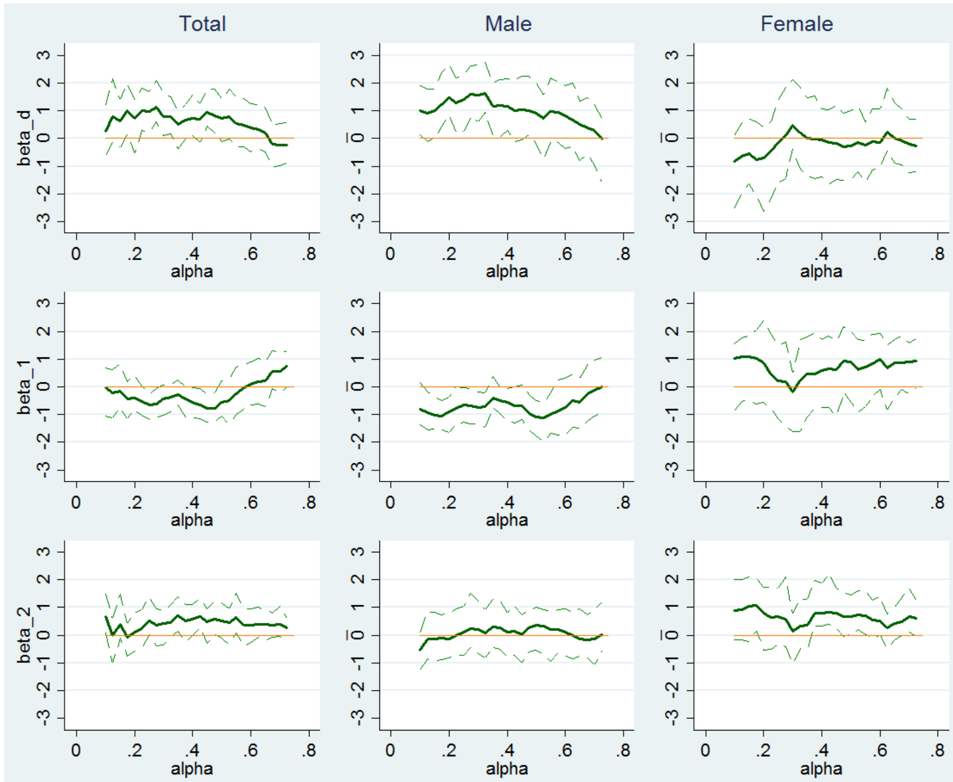

Figure 4: Quantile Effects of UIB at $(c_1, c_2) = (52, 48)$

Turning to Figure 4 for $(c_1, c_2) = (52, 48)$, the results differ much from Figure 3 for $(45, 28)$, because significant positive effects are seen for the full and males only samples while almost no effect for females; the males drive the significant result this time. More specifically, the second column shows that the effects for males are significant over $\alpha = 0.15 \sim 0.4$ with the effect magnitude $1 \sim 1.6$. Since the $0.15 \sim 0.4$ quantile unemployment durations are $2 \sim 4$ months for males, this effect magnitude means that two month additional UIB increases unemployment durations by $2 \sim 6$ months.

As for the partial effects, the partial effect β_1 for males are strongly negative $(-0.4 \sim -1.1)$ and significant over $\alpha = 0.13 \sim 0.55$. If this partial effect is ignored, then the treatment effect would be underestimated for males. On the other hand, the partial effects are insignificant for females. The partial effect β_2 is not significant except for some positive effects for females around the median.

In sum, additional two months of UIB increase unemployment duration differently by cutoff and gender. At the cutoff $(c_1, c_2) = (45, 28)$, mostly females are affected by the treatment; females' unemployment increases by $6 \sim 7$ months. In contrast, mostly males are affected at the cutoff $(c_1, c_2) = (52, 48)$; the effect magnitude is $2 \sim 8$ months. At the age cutoff 52, there are significantly negative partial effects of age for males, which might be related to "pre-retirement". Börsch-Supan and Wilke (2003) mentioned that an unofficial pre-retirement at age 56 was frequent in West Germany because UIB was paid up to three year for those elderly workers. To be eligible for the UIB, they might have avoided unemployment after age 52.

6 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) regression functions other than mean, and (iii) partial effects due to part of the scores crossing cutoff, in addition to the full effect with all scores crossing all cutoffs. The critical difference between our and existing approaches in the literature for multiple-score RD turned out to be partial effects: allowed in this paper, but ruled out in the other papers. We imposed weak continuity assumptions, presented the identified parameters, and proposed simple local difference-in-differences-type estimators for the parameters.

We applied our estimators to German data to find the effects of unemployment insurance benefit (UIB) on ln(unemployment-duration), where UIB requires age and previous working months to cross cutoffs. We found that, at age 45 and working months 28, the α-quantile effects for females are significantly positive ($0.69 \sim 0.84$) over $\alpha = 0.2 \sim 0.6$, and close to zero or insignificant for males. In contrast, at age 52 and working months 48, the effects for males are significantly positive ($1 \sim 1.6$) over $\alpha = 0.15 \sim 0.4$, and close to zero and insignificant for females. When turned into monthly figures, the effect magnitudes mean that two month UIB increase leads to an unemployment duration increase of about $2 \sim 8$ month. We also found substantial partial effects at the latter cutoff, which seems related to an early retirement scheme in Germany.

REFERENCES

Abdulkadiroglu, J. Angrist and P. Pathak, 2014, The elite illusion: achievement effects at Boston and New York exam schools, Econometrica 82, 137-196.

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.

Berk, R., G. Barnes, L. Ahlman and E. Kurtz, 2010, When second best is good enough: a comparison between a true experiment and a regression discontinuity quasi-experiment, Journal of Experimental Criminology 6, 191-208.

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.

Börsch-Supan, A.H. and C.B. Wilke, 2003, The German public pension system: how it was, how it will be, MEA Discussion Paper, 034-03.

Broockman, D.E., 2009. Do congressional candidates have reverse coattails? Evidence from a regression discontinuity design, Political Analysis 17, 418-434.

Caliendo, M., K. Tatsiramos and A. Uhlendorff, 2013, Benefit duration, unemployment duration and job match quality: a regression-discontinuity approach, Journal of Applied Econometrics 28, 604-627.

Calonico, S., M.D. Cattaneo and R. Titiunik, 2014, Robust nonparametric confidence intervals for regression-discontinuity designs, Econometrica 82, 2295-2326.

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.

Chen, M.K. and J.M. Shapiro, 2007, Do harsher prison conditions reduce recidivism? a discontinuity-based approach, American Law and Economics Review 9, 1-29.

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.

Frandsen, B.R., M. Frölich and B. Melly, 2012, Quantile treatment effects in the regression discontinuity design, Journal of Econometrics 168, 382-395.

Hahn, J., P. Todd, and W. Van der Klaauw, 2001, Identification and estimation of treatment effects with a regression discontinuity design, Econometrica 69, 201-209.

Henry, G.T., C.K. Fortner and C.L. Thompson, 2010, Targeted funding for educationally disadvantaged students: a regression discontinuity estimate of the impact on high school student achievement, Educational Evaluation and Policy Analysis 32, 183-204.

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.

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.

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, 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., 2000, Median treatment effect in randomized trials, Journal of the Royal Statistical Society (Series B) 62, 595-604.

Lee, M.J., 2005, Micro-econometrics for policy, program, and treatment effects, Oxford University Press.

Leuven, E., M. Lindahl, H. Oosterbeek and D. Webbink, 2007, The effect of extra funding for disadvantaged pupils on achievement, Review of Economics and Statistics 89, 721-736.

Lipsey, M.W., C. Weiland, H. Yoshikawa, S.J. Wilson and K.G. Hofer, 2015, The prekindergarten age-cutoff regression-discontinuity design: methodological issues and implications for application, Educational Evaluation and Policy Analysis, forthcoming.

Matsudaira, J. D., 2008, Mandatory summer school and student achievement, Journal of Econometrics 142, 829-850.

Michalopoulos, S. and E. Papaioannou, 2014, National institutions and subnational development in Africa, Quarterly Journal of Economics 129, 151-213.

Papay, J.P., R.J. Murnane and J.B. Willett, 2011, Extending the regression discontinuity approach to multiple assignment variables, Journal of Econometrics 161, 203-207.

Powell, J.L., 1986, Censored regression quantiles, Journal of Econometrics 32, 143-155.

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.

Schmieder, J.F., T.V. Wachter and S. Bender, 2012, The effects of extended unemployment insurance over the business cycle: evidence from regression discontinuity estimates over 20 Years, Quarterly Journal of Economics 127, 701-752.

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.

Yao, W. and L. Li, 2014, A new regression model: modal linear regression, Scandinavian Journal of Statistics 41, 656-671.