Institute for Policy Research
Northwestern University
Working Paper Series

WP-10-02

Analyzing Regression-Discontinuity Designs with Multiple Assignment Variables: A Comparative Study of Four Estimation Methods

Vivian C. Wong

Peter M. Steiner

Thomas D. Cook

Institute for Policy Research,
Northwestern University

Version: May 23, 2010

DRAFT
Please do not quote or distribute without permission.

This research was supported in part by grant R305U070003 and R305D100033 from the Institute of Education Sciences, U.S. Department of Education. The authors thank J. R. Lockwood for helpful discussions.

Abstract

In a traditional regression-discontinuity design (RDD), units are assigned to treatment on the basis of a cutoff score and a continuous assignment variable. The treatment effect is measured at a single cutoff location along the assignment variable. A more flexible conceptualization of RDD, however, allows researchers to examine effects along a multi-dimensional frontier using multiple assignment variables and cutoffs. This paper introduces the multivariate regression-discontinuity design (MRDD). For a MRDD with two assignment variables, we show that the overall treatment effect at the cutoff frontier can be decomposed into a weighted average of two univariate RDD effects, and that the weights depend on the scaling of the assignment variables. The paper discusses four methods for estimating MRDD treatment effects—the frontier, centering, univariate, and instrumental variable approaches—and compares their relative performance in a Monte Carlo simulation study under different scenarios. We find that given correct model specifications, all four approaches estimate treatment effects without bias, but the instrumental variable approach has severe limitations in terms of more stringent required assumptions and reduced efficiency.

KEYWORDS: Multivariate regression discontinuity, regression discontinuity, Rubin Causal Model, potential outcomes, instrumental variable.

1. INTRODUCTION

In a traditional regression-discontinuity design (RDD), units are assigned to treatment and comparison conditions solely on the basis of a single cutoff score on a continuous assignment variable. The discontinuity in the functional form of the outcome at the cutoff represents the treatment effect. Proven by Goldberger (2008) in 1972, the design has been shown empirically to produce effects akin to an experiment's (Aiken, West, Schwalm, Carroll, & Hsuing, 1998; Buddelmeyer & Skoufias, 2004; Black, Galdo, & Smith, 2007; Berk, Barnes, Ahlman, & Kurtz, in press; Shadish, Galindo, Wong, Steiner, & Cook, under review). It has been used for program evaluations in criminal justice (Berk & Rauma, 1983), medicine (Finkelstein, Levin, & Robbins, 1996b), economics (see Lee & Lemieux, 2009, for review), and education (Jacob & Lefgren, 2004; Gill, Lockwood, Martorell, Setodji, & Booker, 2007; Wong, Cook, Barnett, & Jung, 2008).

However, units are frequently assigned to treatment on more than one continuous assignment variable. More recent applications of RDD in education have had multiple assignment variables and cutoff scores available for treatment assignment. For example, Jacob and Lefgren (2004) examined the effects of a remedial education intervention that was assigned to students based on missing a reading cutoff, a math cutoff or both. Gill et al. (2007) examined the effects of schools' failure to make Adequate Yearly Progress (AYP) under No Child Left Behind by missing one of 39 possible assignment criteria. Both are examples of the multivariate regression-discontinuity design (MRDD), where treatment effects may be estimated across a multi-dimensional cutoff frontier, as opposed to a single point on the assignment variable. Other MRDD examples in education

research are by Kane (2002), Matsudaira (2008), Papay, Murnane and Willett (in press), and van der Klaauw (2002). With the exception of the latter, all of these studies analyze each of the assignment rules separately. Multiple assignment variables in RDD are not unique to education; they also occur with increasing frequency in other fields of research, such as in the evaluation of labor market programs (Card, Chetty & Weber, 2007; Lalive, Van Ours & Zweimüller, 2006; Lalive, 2008).

Using the potential outcomes notation of the Rubin Causal Model (Holland, 1986; Rubin, 1974), this paper defines the causal estimand τ_{MRD} for a MRDD with two assignment variables (*M* and *R*) and cutoffs. It then shows that the overall treatment effect τ_{MRD} may be decomposed into a weighted average of two univariate RDD effects, τ_M at the *M*-cutoff and τ_R at the *R*-cutoff, and that the weights for τ_{MRD} depend on the scaling of the assignment variables.

The paper also describes four analytic approaches for estimating treatment effects in a MRD design: the frontier, centering, univariate, and instrumental variable (IV) approaches. The frontier approach estimates treatment effects by first modeling the discontinuity at the cutoff frontier using parametric, semiparametric or nonparametric procedures, and then by applying appropriate treatment weights to each cutoff frontier to estimate τ_{MRD}. The frontier approach estimates the overall (τ_{MRD}) and frontier-specific effects (τ_M and τ_R) simultaneously. It is an extension of an approach introduced by Berk and de Leeuw (1999), which relied on parametric regression estimation of the entire response surface under the assumptions of constant treatment effects and a correctly specified regression model. The proposed frontier approach relaxes these assumptions by allowing for heterogeneous treatment effects at the cutoff frontier. In the centering

approach, all assignment variables are centered at their respective cutoffs, and each unit is assigned its minimum centered assignment score. The minimum assignment score is used then as single assignment variable in a traditional univariate RDD. This approach was employed by Gill et al. (2007) in their evaluation of No Child Left Behind. In the univariate approach, researchers choose a single assignment variable and cutoff to estimate an effect, and exclude all observations that are assigned to treatment via the second assignment variable and cutoff. Jacob and Lefgren (2004) applied this approach in their evaluation of Chicago remedial education programs. Finally, in the IV approach, researchers use at least one assignment mechanism as an instrument for treatment receipt and designate units assigned by the second assignment variable and cutoff as treatment-misallocated cases. Although Robinson and Reardon (2009) and Cook et al. (2009) have proposed this approach for analyzing MRDDs, the IV approach has yet to be applied to actual evaluation data.

Though the univariate and centering approaches have been applied in the RD literature, we know of no study that examines systematically the causal estimands and the validity of the four proposed approaches for handling multiple assignment variables. In this paper, we discuss the causal quantities, theoretical underpinnings, and required assumptions for each approach. Through Monte Carlo simulations, we show that the four approaches succeed in yielding unbiased effect estimates when their required assumptions are met. Before introducing the MRDD, we give a brief description of the traditional RDD with a single assignment variable and cutoff (Hahn, Todd, & van der Klaauw, 2001; Trochim, 1984). We then show the MRDD as an extension of the traditional RDD, except that treatment effects are estimated for cutoff frontiers, as

opposed to at a single cutoff point. Although we discuss MRDDs with two assignment variables only, the concepts and analytic approaches presented here extend to MRDDs with more than two assignment variables.

2. THE REGRESSION-DISCONTINUITY DESIGN

2.1 The Regression-Discontinuity Design with a Single Assignment Variable

Under the standard Rubin Causal Model, the causal treatment effect is estimated for a binary treatment intervention D, where $D_i = 0$ if unit i belongs to the control condition and $D_i = 1$ if it belongs to the treatment condition. Let $Y_i(0)$ and $Y_i(1)$ denote the pair of potential outcomes with $Y_i(0)$ as the potential control outcome which is observed if unit i is not exposed to treatment ($D_i = 0$), and $Y_i(1)$ the potential treatment outcome which is observed if unit i is exposed to treatment ($D_i = 1$). In practice, we do not observe both potential outcomes for each unit ("the fundamental problem of causal inference," Holland, 1986). Rather, we observe only the potential treatment or control outcome for each unit depending on the treatment received. Hence, the observed outcome can be written as a function of the potential outcomes: $Y_i = (1 - D_i) \cdot Y_i(0) + D_i \cdot Y_i(1)$.

In a traditional RDD, units are assigned to treatment solely on the basis of a cutoff score (z_c) on a continuous assignment variable (Z). The assignment variable is any measure taken prior to the treatment intervention. Let us assume that a unit gets assigned to the treatment condition if it scores below the cutoff of the assignment variable and to the control condition if its score is equal to or above the assignment variable: $D_i = 1$ if $Z_i < z_c$ and $D_i = 0$ if $Z_i \geq z_c$. When the assignment rule is implemented perfectly, the probability of receiving treatment drops at the cutoff from 1 to 0. More formally, the

discontinuity in the probability of treatment receipt at the cutoff is $\lim_{z \uparrow z_c} E[D_i \mid Z_i = z] - \lim_{z \downarrow z_c} E[D_i \mid Z_i = z] = 1$. Such a RD design is called "sharp" as opposed to a "fuzzy" design (Trochim, 1984) where due to noncompliance, the probability of treatment receipt does not switch from 1 to 0 at the cutoff, but exhibits a jump less than 1: $0 < \lim_{z \uparrow z_c} E[D_i \mid Z_i = z] - \lim_{z \downarrow z_c} E[D_i \mid Z_i = z] < 1$. The jump needs to be strictly positive because a valid RDD requires the treatment probability to be discontinuous at the cutoff (Hahn, Todd, & van der Klaauw, 2001).

In a sharp RDD, the causal quantity of interest is the difference in the potential outcomes at the cutoff, that is,

$$\tau_{SRD} = E[Y_i(1) - Y_i(0) \mid Z_i = z_c] = E[Y_i(1) \mid Z_i = z_c] - E[Y_i(0) \mid Z_i = z_c] \ .$$

Since we never observe control and treatment cases at the cutoff, the causal estimand is better defined in terms of the difference in limits of conditional expectations as we approach the cutoff from below and above:

$$\begin{aligned} \tau_{SRD} &= \lim_{z \uparrow z_c} E[Y_i(1) \mid Z_i = z] - \lim_{z \downarrow z_c} E[Y_i(0) \mid Z_i = z] \\ &= \lim_{z \uparrow z_c} E[Y_i \mid Z_i = z] - \lim_{z \downarrow z_c} E[Y_i \mid Z_i = z] \end{aligned} \tag{1}$$

The second equality is with observed instead of potential outcomes. This holds because we observe only the potential treatment outcomes below the cutoff, and only the potential control outcomes above or at the cutoff. The difference in limits represents the discontinuity (treatment effect) at the cutoff. However, for τ_{SRD} to be interpreted as a causal effect, the potential outcomes must be continuous at the cutoff (Hahn, Todd, & van der Klaauw, 2001; Imbens & Lemieux, 2007; Lee & Lemieux, 2009):

$$\lim_{z \uparrow z_c} E[Y_i(0) \mid Z_i = z] = \lim_{z \downarrow z_c} E[Y_i(0) \mid Z_i = z] \text{ and}$$

$$\lim_{z \uparrow z_c} E[Y_i(1) \mid Z_i = z] = \lim_{z \downarrow z_c} E[Y_i(1) \mid Z_i = z].$$

The causal estimand in a fuzzy RDD, where not all units comply with their treatment assignment, is different from τ_{SRD}. In addition to the continuity assumption for potential outcomes, the fuzzy RDD requires that there are no defiers close to the cutoff (Imbens & Lemieux, 2007; Hahn, Todd, & van der Klaauw, 2001). Defiers are units that take treatment when they are assigned to the control, but enter the control condition when they are assigned to treatment. If the continuity and "no defiers" assumptions are met, then the causal estimand identifies the treatment effect for the subpopulation of compliers at the cutoff (Imbens & Lemieux, 2007; Hahn, Todd, & van der Klaauw, 2001), where unit *i* is a complier (*C*) if it adheres with treatment assignment: $i \in C$ if $\lim_{z \uparrow z_i} D_i(z) = 1$ and $\lim_{z \downarrow z_i} D_i(z) = 0$, with *D* being the treatment indicator as defined above. Here, unit *i* takes the treatment when assigned to treatment and it takes the control condition when assigned to the control. The causal estimand for the fuzzy RDD is given by $\tau_{FRD} = E[Y_i(1) - Y_i(0) \mid Z_i = z_c, i \in C]$ or in limits notation by

$$\begin{aligned} \tau_{FRD} &= \lim_{z \uparrow z_c} E[Y_i(1) \mid Z_i = z_c, i \in C] - \lim_{z \downarrow z_c} E[Y_i(0) \mid Z_i = z_c, i \in C] \\ &= \frac{\lim_{z \uparrow z_c} E[Y_i \mid Z_i = z_c] - \lim_{z \downarrow z_c} E[Y_i \mid Z_i = z_c]}{\lim_{z \uparrow z_c} E[D_i \mid Z_i = z_c] - \lim_{z \downarrow z_c} E[D_i \mid Z_i = z_c]}. \end{aligned} \tag{2}$$

2.2 The Multivariate Regression-Discontinuity Design with Two Assignment Variables

The MRDD has an assignment process that is based on two or more assignment variables. In this paper, we consider only sharp MRDDs with two assignment variables, *R* and *M* with respective cutoffs r_c and m_c. Units are assigned to treatment if they miss

cutoff r_c, m_c, or both. Figure 1 shows that units are assigned to the control condition *C* if they score above both cutoffs ($R_i \geq r_c, M_i \geq m_c$) and to the treatment condition *T* if they score below either cutoff ($R_i < r_c$ or $M_i < m_c$). We partition the treatment assignment space into three subsets: T_1 if units miss only cutoff r_c, T_3 if they miss only cutoff m_c, and T_2 if they miss both cutoffs. Though 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 this design, *R* and *M* may be reading and math test scores (respectively), treatment may be a standardized test preparation course, and assignment to treatment may be based on whether students fail to achieve minimum threshold scores for reading or math. Although this is a fairly specific implementation of a MRDD, the results presented here may also apply to MRDDs where treatment and control conditions are swapped. Figure 1 shows the cutoff frontier $F = \{(r,m) : (r \geq r_c, m = m_c) \cup (r = r_c, m \geq m_c)\}$ at which the average treatment effect is estimated. Assuming complete treatment compliance, the average treatment effect at the cutoff frontier is given by

$$\tau_{MRD} = E[Y_i(1) - Y_i(0) \mid (R_i, M_i) \in F]. \quad (3)$$

Decomposition of the average treatment effect τ_{MRD}. Since the cutoff frontier consists of the *R*-frontier along assignment variable *M*, $F_R = \{(r,m) : (r = r_c, m \geq m_c)\}$, and the *M*-frontier along assignment variable *R*, $F_M = \{(r,m) : (r \geq r_c, m = m_c)\}$, we can decompose the treatment effect into a weighted average of the treatment effects at the *R*- and *M*-frontiers. Let the difference in potential outcomes be $G_i = Y_i(1) - Y_i(0)$ and the joint density function for assignment variables *R* and *M* be $f(r,m)$, then, we can define

the average treatment effect at the cutoff frontier F as the weighted average of conditional expectations given the single frontiers F_R and F_M (see Appendix A for the proof):

$$\begin{aligned}\tau_{MRD} &= E[G_i \mid (R_i, M_i) \in F] = w_R E[G_i \mid R_i \in F_R] + w_M E[G_i \mid M_i \in F_M] \\ &= w_R \tau_R + w_M \tau_M \ ,\end{aligned} \tag{4}$$

where weights w_R and w_M reflect the probabilities for observing a unit at the R- or M-frontier,

$$w_R = \frac{\int\limits_{m \geq m_c} f(r = r_c, m)dm}{\int\limits_{m \geq m_c} f(r = r_c, m)dm + \int\limits_{r \geq r_c} f(r, m = m_c)dr} \quad \text{and} \tag{5}$$

$$w_M = \frac{\int\limits_{r \geq r_c} f(r, m = m_c)dr}{\int\limits_{m \geq m_c} f(r = r_c, m)dm + \int\limits_{r \geq r_c} f(r, m = m_c)dr} .$$

The conditional expectations represent the average treatment effects τ_R and τ_M at the two discontinuity frontiers F_R and F_M since

$$\tau_R = E[G_i \mid R_i \in F_R] = \frac{\int\limits_{m \geq m_c} g(r, m) f(r = r_c, m)dm}{\int\limits_{m \geq m_c} f(r = r_c, m)dm} \quad \text{and} \tag{6}$$

$$\tau_M = E[G_i \mid M_i \in F_M] = \frac{\int\limits_{r \geq r_c} g(r, m) f(r, m = m_c)dr}{\int\limits_{r \geq r_c} f(r, m = m_c)dr} ,$$

where $g(r, m) = y_1(r, m) - y_0(r, m)$ is the difference in potential outcomes. As shown in Appendix A, we may also use conditional and marginal distributions for defining weights and conditional expectations which is more convenient for estimating the treatment effects.

Assumptions required for MRDD. Given that τ_{MRD} may be decomposed into a weighted average of frontier-specific effects, all required assumptions for the traditional univariate RDD must be met for each discontinuity frontier (F_R and F_M). First, the design requires a discontinuity in treatment probabilities at F_R and F_M. Second, the expectations of potential outcomes need to be continuous at F_R and F_M. Third, for fuzzy MRDDs, no defiers are allowed close to and at F_R and F_M.

The second assumption requires some elaboration. For the potential treatment outcomes, the continuity assumption states that the limits of the expected values have to be identical at the cutoff frontiers:

$$\lim_{r \uparrow r_c} E[Y_i(1) \mid R_i = r, M_i \geq m_c] = \lim_{r \downarrow r_c} E[Y_i(1) \mid R_i = r, M_i \geq m_c] \text{ and}$$
$$\lim_{m \uparrow m_c} E[Y_i(1) \mid R_i \geq r_c, M_i = m] = \lim_{m \downarrow m_c} E[Y_i(1) \mid R_i \geq r_c, M_i = m] \ .$$

The same equality must hold for potential control outcomes $Y(0)$. Two important remarks need to be made. First, no continuity is required for $M < m_c$ at $R = r_c$ and for $R < r_c$ at $M = m_c$ because these frontiers do not belong to the discontinuity frontier F (in Figure 1, these are the dashed frontiers between treatment subsets T_1 and T_2, and T_2 and T_3). Second, the continuity assumption does not imply continuity of $E[Y(1)]$ and $E[Y(0)]$ along frontier F. Consider an arbitrary point $r^* \geq r_c$ at cutoff frontier F_M. Then, $\lim_{r \uparrow r^*} E[Y_i(1) \mid R_i = r, M_i = m_c]$ may differ from $\lim_{r \downarrow r^*} E[Y_i(1) \mid R_i = r, M_i = m_c]$. In particular, the expectation of potential outcomes may be discontinuous at the intersection point $(r = r_c, m = m_c)$ where frontiers F_R and F_M meet: $\lim_{m \downarrow m_c} E[Y_i(1) \mid R_i = r_c, M_i = m]$ may differ from $\lim_{r \downarrow r_c} E[Y_i(1) \mid R_i = r, M_i = m_c]$, and both may differ from $E[Y_i(1) \mid R_i = r_c, M_i = m_c]$ at the intersection point. Such a discontinuity at the intersection point has no impact on the

frontier-specific treatment effects τ_M and τ_R because the intersection point has a probability mass of zero with respect to the discontinuity frontiers F_M and F_R. The same holds for the potential control outcome $Y(0)$.

Nonetheless, continuity in expectations of both potential outcomes along frontier F, particularly at the intersection point (r_c, m_c), is desirable in practice as Figure 2 illustrates (for clarity we only show the response surface of $E[Y(1)]$). The left panel shows a continuous response surface of $E[Y(1)]$ along the cutoff frontier F (solid line) but also along the entire cutoffs r_c and m_c (i.e., including the dashed lines). The right panel illustrates a case where $E[Y(1)]$ is discontinuous at the intersection point, but continuous at each of the two frontiers F_M and F_R. This requires a rather awkward functional form with several discontinuities (e.g., a discontinuity along the diagonal for the control cases) to ensure that the potential treatment outcomes are connected smoothly at both frontiers. The response surface of the potential treatment outcome presented here seems rather implausible with a single treatment. However, with multiple treatments (where a unique treatment is assigned to each subspace T_1, T_2, and T_3), such discontinuities in the response surface of the potential treatment outcomes would be plausible, if not expected (the formalization would require four potential outcomes and a careful definition of the causal estimands of interest).

Scale-dependency of the average treatment effect τ_{MRD}. The decomposition of the average treatment effect of a MRDD into a weighted average of unviariate RDD effects, τ_R and τ_M, reveals that the average treatment effect τ_{MRD} depends on weights w_R and w_M. Since the weights are determined by integrating the joint density $f(r,m)$ along frontier F, their ratios depend crucially on the scaling of assignment variables R and M.

For instance, the weight w_R for the treatment effect at the *R*-frontier decreases relative to w_M as assignment variable *R* is rescaled to $R_s = sR$ with $s > 1$ (with scaling of *M* held constant). Conversely, w_R's relative weight increases when the *R* assignment variable is rescaled with $s < 1$. This is because of the disproportional change in integrals $\int_{m>m_c} f(r = r_c, m)\,dm$ and $\int_{r>r_c} f(r, m = m_c)\,dr$ in equation (5) when *R* and *M* are rescaled with different scaling factors *s*. Figure 3 shows that when *R* is rescaled into $R_s = 2R$, such that the variance of the rescaled variable is four times larger than the original variable's variance, the ratio of the weights also changes (as indicated by the shaded areas along the cutoff frontier). Since the average treatment effect τ_{MRD} is sensitive to the choice of the assignment variables' scale, an infinite number of average treatment effects exist. This is an unpleasant property of MRDD that is of special relevance whenever the average treatment effects for treatment frontiers F_M and F_R differ ($\tau_M \neq \tau_R$). We discuss practical implications of the scale-dependency in the discussion section.

3. ESTIMATION STRATEGIES FOR MRDD

A MRDD with two assignment variables allows the estimation of three different causal quantities: two frontier-specific effects τ_R and τ_M and an overall effect τ_{MRD}. In this paper, we present the following four estimation procedures: the frontier, centering, univariate, and instrumental variable approaches. The first two approaches aim at estimating the overall treatment effect τ_{MRD}, and the latter two at the frontier-specific effects τ_R and τ_M .

Frontier approach. This procedure estimates the discontinuity along both frontiers simultaneously, and applies appropriate weights to obtain the overall effect τ_{MRD} . It first estimates the discontinuous response surface $\hat{y}(r,m)$ using a parametric, semi-, or nonparametric regression method. Since we are interested only in the treatment effect at the cutoff frontier *F,* the treatment function $\hat{g}(r,m \mid F)$ is estimated by taking the difference in the estimated treatment outcome $\hat{y}_1$ and the control outcome $\hat{y}_0$ along F such that $\hat{g}(r,m \mid F) = \hat{y}_1(r,m \mid F) - \hat{y}_0(r,m \mid F)$. Then, the joint density function $\hat{f}(r,m)$ is estimated by using a bivariate kernel-density estimator. Finally, we plug $\hat{g}(r,m \mid F)$ and $\hat{f}(r,m)$ into equations (5) and (6) and estimate τ_R and τ_M , as well as weights w_M and w_R (for estimating τ_{MRD}), by numerical integration.

Since an accurate estimation of the bivariate density requires intense computational resources and a large number of observations, a nonparametric estimation of the univariate conditional and marginal densities $f_{R|M}$, $f_{M|R}$, f_R , and f_M is preferable. Thus, by plugging the estimated univariate densities and treatment function into equations (A3) and (A4) and by using numerical integration, we obtain frontier-specific effects $\hat{\tau}_R$ and $\hat{\tau}_M$, weights $\hat{w}_R$ and $\hat{w}_M$, and the average treatment effect $\hat{\tau}_{MRD}$. This strategy also works with smaller datasets and for MRDDs with more than two assignment variables since univariate kernels suffice for estimating conditional and marginal densities. Because $\hat{f}(r,m)$ and marginal densities $\hat{f}_R$ and $\hat{f}_M$ depend on the scaling of *R* and *M*, the choice of different bandwidths at the *R*- and *M*-cutoffs also would affect the ratio of weights. As a result, we recommend using the same bandwidth (in absolute units) for both dimensions. Bootstrapping may be used for estimating standard

errors. However, the nonparametric estimation of densities and numerical integration is cumbersome, data-hungry and computationally expensive (particularly in bootstrapping standard errors). The centering approach, which we discuss next, tries to overcome these issues by downscaling the multiple assignment variables into a single composite assignment variable.

Centering approach. This procedure collapses multiple assignment scores into a single assignment variable, thereby reducing a high-dimensional assignment mechanism to a one-dimensional mechanism. This is achieved by the following procedure: For each assignment variable, center each unit's score at its respective cutoff, such that $R_i^z = R_i - r_c$ and $M_i^z = M_i - m_c$. Then, choose the minimum centered value as the unit's sole assignment score: $Z_i = \min(R_i^z, M_i^z)$. The minimum applies only in MRD designs where the top right quadrant of the assignment variable plane is the control condition (quadrant *C* in Figure 1). In cases where a different segment of the surface is the control region (e.g. quadrant T_2), sign-transformations of the assignment variables or choosing the *maximum* centered assignment score is required for creating the composite assignment variable. Finally, apply standard RD analytic methods (e.g. local polynomial regression) for estimating treatment effects by using *Z* as the assignment variable and zero as the cutoff.

Despite these dimension-reducing transformations, the centering approach estimates the same causal estimand as defined in equation (3). First note that at cutoff z_c where $\min(R_i^z, M_i^z) = 0$, the population consists of two subpopulations: subjects from the *R*-frontier $F_R = \{(r,m) : r - r_c = 0, m - m_c \geq 0\}$ and subjects from the *M*-frontier $F_M = \{(r,m) : r - r_c \geq 0, m - m_c = 0\}$. Then, the treatment effect can be decomposed

further into the weighted average of frontier-specific treatment effects as defined in equation (4):

$$
\begin{aligned}
E[G_i \mid Z_i = 0] &= E[G_i \mid \min(R_i^z, M_i^z) = 0] \\
&= E[G_i \mid (R_i^z = 0, M_i^z \geq 0) \text{ or } (R_i^z \geq 0, M_i^z = 0)] \\
&= E[G_i \mid (R_i, M_i) \in F_R \cup F_M] = w_R E[G_i \mid R_i \in F_R] + w_M E[G_i \mid M_i \in F_M]
\end{aligned}
$$

where weights and conditional expectations are given as before (equations (5) and (6)). This result implies that the causal quantity estimated by the centering approach is also sensitive to the scaling of assignment variables. For example, increasing the scale for assignment variable *R* moves observations for which $\min(R_i^z, M_i^z) = R_i^z$ is farther away from the cutoff $z_c = 0$. Thus, units assigned by the *R* assignment mechanism would receive less weight relative to observations with $\min(R_i^z, M_i^z) = M_i^z$.

The chief advantage of the centering approach is that it allows the researcher to collapse scores from multiple assignment rules to a single assignment variable. The approach also generalizes well to MRDDs with more than two assignment variables, as well as simplifies the analyses for estimating average treatment effects across multiple discontinuity frontiers. However, it does not allow the estimation of frontier-specific effects $\hat{\tau}_R$ and $\hat{\tau}_M$ but that can be done with one of the two approaches described next.

Univariate approach. This approach solves the dimensionality problem by estimating treatment effects for each frontier separately. We estimate the treatment effect at the *R*-frontier F_R by excluding all observations scoring on assignment variable *M* below the cutoff m_c since the cutoff frontier F_R is defined only for $M \geq m_c$. Then, using standard RDD methods like local polynomial regression (Imbens & Lemieux, 2008), we estimate the treatment effect at F_R according to equation (1). We estimate the treatment

effect at the cutoff frontier F_M in a similar way, except that we exclude observations scoring below the cutoff on assignment variable *R*. Also, the average treatment effect τ_{MRD} may be estimated, but it requires calculation of appropriate treatment weights for each frontier as described for the frontier approach.

Instrumental variable approach. Rather than excluding observations assigned to treatment by alternative mechanisms, we estimate frontier-specific effects by delegating some of these cases as "fuzzy" units. To estimate the treatment effect at the *R*-cutoff r_c, we include all units below and above the cutoff, but designate units for which $R_i \geq r_c$ and $M_i < m_c$ as fuzzy cases (these are units in quadrant T_3 in Figure 1). Thus, the instrument for treatment receipt is derived from the *R* assignment variable and cutoff alone. The local average treatment effect at r_c is, according to equation (2), given by the ratio of the difference in the new treatment and control group's mean values and the difference in compliance rates at the cutoff. Nonparametric methods are typically used for estimating the complier average treatment effect at the cutoff (Imbens & Lemieux, 2008).

By focusing on the unrestricted cutoff frontier at r_c instead of the restricted cutoff frontier F_R, the IV approach estimates a different causal quantity from what is implied by the MRDD. The IV estimates the local average treatment effect along the entire length of the *R*-cutoff, while the univariate RD approach estimates the design induced average treatment effect along the *R*-cutoff frontier F_R only. However, the two causal quantities are identical if the average treatment effects for both subfrontiers, $(r = r_c, m \geq m_c)$ and $(r = r_c, m < m_c)$, are identical, but the continuity of expected potential outcomes must hold for each subfrontier separately. In the sharp MRDD case, the IV approach implicitly

assumes the equivalence of the two treatment effects since all units in quadrant T_3 are non-compliers. Hence, the treatment effect for subfrontier $(r = r_c, m < m_c)$ cannot be estimated because all units in T_2 and T_3 are treated units, forcing the IV to infer the treatment effect only from subfrontier $(r = r_c, m \geq m_c)$ and to assume the same effect for subfrontier $(r = r_c, m < m_c)$.

4. SIMULATION DESIGN FOR MONTE CARLO COMPARISON OF APPROACHES

In a series of simulation studies, we examine the performance of the frontier, centering, univariate, and IV approaches when we vary the following three factors: 1) complexity of the "true" response surface; 2) distribution and scale of the assignment variables; and 3) approach for analyzing a sharp MRDD.

The first factor varies the complexity of the "true" response surface. The goal here is to assess how each approach performs when the "true" response surface is straightforward to model, versus when it is complex with heterogeneous treatment effects. The outcome Y_i for unit i is a simulated math score based on three different specifications of the true response surface:

Model 1: $Y_i = 4T_{0i} + .5R_i + 1M_i + \varepsilon_i$

Model 2: $Y_i = 4T_{0i} + .5R_i + 1M_i - .05T_{1i}M_i + .55T_{3i}R_i - .025T_{1i}R_iM_i - .005T_{3i}R_iM_i + \varepsilon_i$

Model 3: $Y_i = 4T_{0i} - 2T_{1i} + 2T_{3i} + .5R_i + 1M_i - .05T_{1i}M_i + .55T_{3i}R_i - .025T_{1i}R_iM_i - .005T_{3i}R_iM_i + \varepsilon_i$

where R_i and M_i are the assignment variables, drawn from a bivariate normal distribution with a correlation of 0.2. T_i equals 1 if unit i receives any treatment at all and 0 if it did not, and as depicted in Figure 1, T_{1i} equals 1 if unit i has an R assignment score less than

r_c but an M assignment score greater than m_c (and 0 if otherwise) and T_{3i} equals 1 if unit i has an M assignment score less than m_c but an R assignment score greater than r_c (and 0 if otherwise). ε_i is a normally distributed error term with a mean of zero and a standard deviation of 2. Model 1 ("constant treatment effects model") shows a constant treatment effect of 4.00, with no treatment by assignment variable interactions and level changes in the treatment response surface (Model 1 in Figure 4). Model 2 ("heterogeneous treatment effects model") defines heterogeneous effects based on continuous potential treatment outcomes. In this model, effects along F_R and F_M are heterogeneous due to interactions between the treatment and assignment variables (Model 2 in Figure 4). Our third model ("heterogeneous but discontinuous effects model") defines heterogeneous effects based on discontinuous potential treatment outcomes. The model indicates heterogeneous treatment effects across the cutoff frontiers, as well as level changes in the treatment response surface between T_1 and T_2 and between T_2 and T_3 (Model 3 in Figure 4). Level changes in the response surface might occur if the treatment condition for T_1 and T_3 vary greatly from the treatment condition for T_2. As discussed above, we expect that this model will produce biased effect estimates for the IV approach.

Given the above data-generating equations and the distribution of assignment variables, we first computed for each model the true treatment effects for frontiers F_R and F_M and F. These theoretical effects serve as our benchmarks for comparing the estimates produced by the four approaches. Since the overall treatment effect τ_{MRD} depends on the scaling of assignment variables, we present two "true" effects for each model: one for the raw, unstandardized assignment variables and one for the standardized assignment variables.

The second factor we vary examines how differences in the distribution and scale of the two assignment variables in a MRD design affect the performances of the proposed approaches. First, we look at two assignment variables that are on the same scale with identical distributional shapes. Both variables *R* and *M* are normally distributed with the same standard deviation of 10, but have different means and cutoffs. For the *R* assignment variable, the mean is 45 and the cutoff is 40. For the *M* assignment variable, the mean is 55 and the cutoff is 60. Second, we examine a MRDD with assignment variables on the same scales as above, but with different standard deviations. In this scenario, *R* has a standard deviation of five and *M* has a standard deviation of 20. Finally, we look at a MRDD with assignment variables on different scales, where *M* has the same mean, distribution, and cutoff as in the first scenario, but we transformed the *R* assignment variable such that $R_s = R/100$. Thus, the new *R* cutoff is .40, the mean is .45, and the standard deviation is .10.

The third factor defines the four methodological procedures that we study. For the frontier approach, we use two parametric specifications of the regression models: the full model, which includes all covariates and interaction terms as defined in Model 3, and the constant treatment effects model, which assumes constant treatment effects across the response surface as in Model 1. This constant treatment effects model is equivalent to Berk and de Leeuw's suggestion (1999).[1] The goal is to assess the degree of bias when the treatment function is misspecified. We estimate the response surface via parametric regression and the conditional and marginal densities at the *R*- and *M*-cutoffs using a

[1] In using parametric regression, we could have estimated the overall and frontier-specific effects directly without integrating along the frontier. However, a semi- or nonparametric estimation of the response surface would require integration for estimating average treatment effects.

kernel density estimator with an Epanechnikov kernel. Finally, we numerically integrate the product of treatment and density functions along the cutoff frontiers to obtain the conditional expectation across both frontiers.

For the centering, univariate and instrumental variable approaches, we estimate treatment effects using local linear kernel regression (Imbens & Lemieux, 2008). In each iteration of the simulation, the bandwidths are selected based on Imbens and Kalyanaraman's (2010) algorithm for optimal bandwidth choice at the cutoff. While the centering approach only allows the estimation of the overall treatment effect τ_{MRD}, the univariate and instrumental variable approaches focus on the frontier-specific treatment effects τ_M and τ_R. For the centering approach, we estimate treatment effects using both the raw and standardized scores because recent applications of the centering approach (e.g., Gill et al., 2007) use standardized assignment variables for estimating treatment effects.

The goal of the Monte Carlo study with 500 simulated samples of size 5,000 is to evaluate the unbiasedness of the proposed approaches for estimating the true treatment effects. Due to computational reasons, we do not directly investigate variance estimators of the treatment effects (for each approach, standard errors may be bootstrapped). However, for assessing the relative efficiency of the four approaches, we report standard errors estimated from our simulation, which is the standard deviation of estimated treatment effects across the 500 iterations, $s_\tau = \sqrt{\sum_{i=1}^{500}(\hat{\tau}_i - \bar{\tau})^2}$, where $\bar{\tau}$ is the average of the estimated effects $\hat{\tau}_i$ (the simulation standard deviation may be considered as an average standard error bootstrapped from the overall target population instead of the actual samples). To test the unbiasedness of an estimated treatment effect with respect to

its corresponding true effect, we use simulation standard errors $s_\tau / \sqrt{500}$. Significant differences between estimated treatment effects and true effects (at the .05 error level) are indicated by asterisks in the tables (simulation standard errors are not presented in the tables).

5. MONTE CARLO RESULTS

5.1 Effect Estimates for MRDDs with Constant Treatment Effects

The first line of the column panels in Table 1 presents the theoretical treatment effects (τ_M , τ_R, and τ_{MRD}) according to the data-generating model (see Model 1, Figure 4). The second line shows the theoretical treatment effect when both assignment variables are standardized before analyzing the MRDD. Regardless of the distribution and scale of the assignment variables, all the true effects are 4.00 when the raw or standardized assignment scores are used. This demonstrates that standardizing assignment scores does not affect the causal quantities estimated for the frontier-specific and overall effects whenever treatment effects are constant. Note, however, that weights w_M and w_R remain sensitive to the scaling and distribution of the assignment variables.

The first panel of Table 1 shows the results for two assignment variables on the same scale with identical distributions (first panel of Table 1). The frontier approach results in treatment effects between 3.99 and 4.01 at the F_R and F_M frontiers, and 4.00 for the overall effect. Of the four approaches examined, the frontier approach yields the most precise estimates. Treatment weights produced by the frontier approaches are also comparable in terms of relative proportion to the theoretical weights. The centering approach produces an overall effect of 3.98 when raw assignment scores are used and an

effect of 3.94 when standardized scores are applied. The latter is significantly different from the theoretical effect, suggesting that it is slightly biased. Treatment standard errors for both centering estimates are approximately four times larger than those generated by the frontier approach. The univariate approach replicates treatment effects for both frontiers, but again, standard errors are between five (F_M) and 10 times (F_R) larger than those produced by the frontier approach. Although the instrumental variable approach estimates effects that are not significantly different from the theoretical true effects, they are less efficient. The treatment standard error is .81 for τ_M and 3.64 for τ_R (the standard errors differ considerably due to the differential strength of instruments).

The second column panel of Table 1 shows weights and treatment effects when assignment variables are on the same scale but have different distributions (i.e., standard deviations). Except for the centering approach with unstandardized assignment variables, all methods replicate the theoretical treatment results. The frontier approach yields treatment estimates that range from 3.99 to 4.02, with treatment standard errors that are between .01 and .16. The centering approach yields an overall effect of 3.86 when the raw assignment score is used, and an effect of 4.00 when standardized scores are applied. The treatment effect for the raw assignment score (3.86), however, slightly underestimates the theoretical effect of 4.00. The univariate and IV approaches produce effects that are not significantly different from their theoretical effects, but they are less efficient than those generated by the frontier approach.

We also examined the performance of the proposed approaches when the MRDD is based on two assignment variables on different scales (third column panel of Table 1). As before, all procedures replicate the theoretical effects of 4.00 points, except for the

centering approach with standardized assignment variables. Estimates obtained by the frontier approach are the most efficient, while the IV estimates are the least efficient.

Overall, the frontier, univariate and IV approaches succeed in replicating the theoretical estimates when treatment effects are constant. The exception is the centering approach, which produced biased results half the time. The significant, but small bias, is a result of pooling units from multiple cutoff frontiers into a single composite variable. This is because of two reasons (given non-constant response surfaces for potential outcomes): First, pooling units from different frontiers increases the heterogeneity of the outcome at the pooled cutoff, requiring a larger bandwidth for nonparametric estimates. The larger bandwidth, however, might introduce bias in a local polynomial regression due to mis-specification of the functional form. Second, pooling increases the complexity of the functional form around the cutoff. Even in the simple case with linear response surfaces and a constant treatment effect (Model 1), pooling produces a nonlinear relation between the composite assignment variable and the outcome due to differential distributions of frontier-specific units in the neighborhood of the single cutoff. Modeling a quadratic or cubic polynomial in the local regression would mitigate the bias, but it would also reduce the efficiency of the estimates. This result highlights the sensitivity of the centering approach to mis-specification in the response function, even when nonparametric regression methods are used. The univariate and IV approaches, which also use local linear regression for estimating effects, do not exhibit the same bias because the local linearity assumption holds for these procedures.

5.2 Estimates for MRDDs with Heterogeneous Treatment Effects

Table 2 presents results for a MRDD with heterogeneous treatment effects across the treatment response surface (see Model 2, Figure 4). When the assignment variables are on the same scale with similar distributions, the effect for F_M and F_R are 9.74 and 3.70, and for the overall effect, it is 8.11 points (first two lines of the first panel in Table 2). Note that the overall effect is the same regardless of whether the standardized or raw assignment scores were used, even when treatment effects are heterogeneous. When the MRDD consists of two assignment variables on the same scale but have different distributions, theoretical treatment estimates are 7.59 points for F_M and 3.36 points for F_R, for both the raw and standardized effects (second panel). However, while the frontier-specific effects do not depend on scaling, the overall treatment effect does. When raw assignment scores are used, the overall theoretical effect is 5.54, but when standardized scores are used, the theoretical effect is 6.79 points. The differences in these two effects show that standardizing the assignment variables changes the relative proportion of the treatment weights when a MRDD with heterogeneous effects has two assignment variables with different distributions. Standardizing changes the weight ratio of frontiers F_M and F_R from 51.5:48.5 to 80.9:19.1 (Table 2 shows the weights in percentage terms that sum to 100). Finally, for a MRDD with assignment variables on different scales and with heterogeneous treatment effects (third column panel of Table 2), the true overall treatment effect also depends on the scaling due to the differences in weight ratios.

Table 2 shows that all four methods generally perform as expected when effects are heterogeneous along the cutoff frontiers. The frontier approach produces unbiased effects when the response function is correctly specified (full model), with the only

exception being an estimate for F_M (first panel of Table 2). However, this exception is caused by the slight bias in the frontier weights, which is mostly likely due to chance because the frontier approach produces no other significant differences whenever the model is correctly specified. When the treatment functions are incorrectly modeled, the frontier approach performs poorly in reproducing the theoretical effects. The constant treatment effects model—which assumes constant effects across the treatment response surface—yields estimates that are significantly different from their benchmark effects (for raw assignment variables). This is true regardless of whether the assignment variables are on the same scale and distributions, on the same scales with different distributions, or when the assignment scores are on different scales. While the centering approach produces biased results in two of the six estimates (for the same reason as discussed above), the univariate and IV approaches produce unbiased effect estimates in every case. Also note that the centering approach with standardized assignment variables estimates the corresponding true effect for standardized scores (which differs from the true effect for raw scores).

5.3 Estimates for MRDDs with Heterogeneous Treatment Effects but Discontinuous Potential Treatment Outcomes

Table 3 presents results for MRDDs with heterogeneous treatment effects but with discontinuous potential treatment outcomes (see Model 3, Figure 4). When the assignment variables are on the same scale with identical distributions, the true treatment effect for F_M is 8.61 points, and for F_R, it is 1.70 points, making the overall effect across both frontiers 6.74 points (first column panel). These effects are identical for when the

raw and standardized assignment scores are used, showing again that when the assignment variables are on the same scale with identical distributions, standardizing does not change the relative proportions of the treatment weights. If the two assignment variables differ in their distributions, the true overall treatment effect for the raw assignment variables is 4.59 points but 6.44 points when assignment variables are standardized. The difference between the raw and standardized theoretical effect is reflected in the frontier weights, 51.5 and 48.5 for F_M and F_R, respectively, when raw scores are used, and 80.9 and 19.1 when standardized scores are applied. Similarly, for assignment variables on different scales, the overall effect depends on the scaling: 8.59 for the raw scores and 6.82 for the standardized scores (third panel of Table 3).

Simulation results presented in Table 3 show that when effects are heterogeneous but with discontinuities in potential treatment outcomes, the IV approach fails to generate unbiased effect estimates for either frontier. The biases are large and significantly different from the theoretical benchmarks. The centering approach continues to produce mixed results, with small biases for a third of the estimates. The frontier approach generally performs well if the treatment function is correctly specified, and the univariate approach does well when the bandwidths and the degree of the polynomial are correctly specified for local nonparametric estimates.

6. DISCUSSION

What do results presented in this paper imply for practice? The obvious question for most researchers is, "Which approach should I use for estimating treatment effects in a MRDD?" In comparing the relative benefits and limitations of the four approaches, our

recommendation is to start with analyzing each frontier of the MRDD separately by using the univariate approach. Results from our simulation study indicate that the nonparametric univariate approach performed well in estimating frontier-specific effects, with no significant differences between the estimated and theoretical effects for F_M and F_R, given a reasonable choice of the local polynomials degree and bandwidths.

The researcher can then assess whether treatment effects are constant across both cutoff frontiers, and if so, use the frontier or centering approach to estimate an overall effect τ_{MRD}. In general, the frontier approach performs well in estimating τ_{MRD} when the treatment functions are correctly specified. It also has the advantage of improved statistical efficiency for both the overall and frontier-specific effects. However, because the true functional from of the response surface is hardly ever known in practice, researchers may consider using the centering approach for estimating τ_{MRD} to avoid the complication of modeling a multi-dimensional response surface. The issue here is that the centering approach may be prone to small biases due to wider bandwidths and more complex functional forms caused by pooling units from different cutoff frontiers. However, the bias might be mitigated by using difference scores as the outcome whenever pretests are available to reduce the complexity of the functional form and heterogeneity of the dependent variable.

In general, we recommend against using the IV approach for estimating frontier-specific effects τ_M and τ_R. Although it yields unbiased estimates when its analytic assumptions are met, simulation results indicate that the IV approach has reduced statistical precision as compared to the other three methods. The approach also yields biased results when there are discontinuities along the extended cutoff frontiers in the

potential treatment outcomes. A theoretical concern with the IV approach is that the causal quantity implied by the MRDD differs from the causal quantity estimated by the instrumental variable. Taken together, the IV approach offers no comparative advantages over the other three proposed methods examined in this paper.

A second practical consideration is whether to use standardized or raw assignment scores for the centering and frontier approaches. Theory and simulation results indicate that the weighted average treatment (τ_{MRD}) is sensitive to the scaling of the assignment variables. When assignment variables are transformed (e.g. into standardized values), the causal quantity of the overall effect changes. For example, consider two possible assignment variables, household income and age. Here, treatment weights depend heavily on the units of measurement for each assignment variable, where a measure of income in thousands of dollars instead of dollars would drastically increase the weight of the "income frontier" in the overall treatment effect. Standardizing the income and age variables, such that both have a standard deviation of one, does not solve the scaling issue because the procedure fails to provide a substantive interpretation of the overall effect. Standardizing is only one possibility for transforming assignment scores into a common metric—another option is to use rank order scores. But all these methods result in different weights for the treatment frontiers and, thus, different causal quantities.

A second concern with standardizing assignment variables is that the procedure is sensitive to distribution properties that would have strong effects on the standard deviation used for scaling. For example, the presence of extreme observations for one assignment variable would result in a down-weighting of the treatment effect for the corresponding frontier (due to the large standard deviation). The only case when scaling

is not an issue for estimating τ_{MRD} is when the average treatment effects for both frontiers are identical. However, we believe this scenario to be rare in practice. In any case, τ_M and τ_R provide at least informative upper and lower bounds for the overall treatment effect.

In this paper, we have assumed that units are assigned to a single treatment condition via two assignment variables and cutoffs, but this need not be the case. Units may be assigned to one treatment condition for missing the *R* cutoff, another if they miss the *M* cutoff, and both if they miss the *R* and *M* cutoffs together. This type of MRD design raises several practical considerations for the researcher. First, although the frontier and univariate assignment variable approaches can estimate unbiased frontier-specific effects (τ_M and τ_R), the IV approach may yield biased results due to violations in the continuity assumption because variations in treatment conditions may introduce discontinuities in potential treatment outcomes. Second, a MRDD with multiple treatment conditions raises questions about whether estimating an overall effect is appropriate and substantively interpretable given that it averages estimates across two unique treatments. Third, treatment contrasts in a MRD design are limited to comparisons along the cutoff frontiers. There may be cases, however, when the desired treatment contrast is to compare outcomes from units that received the "strongest" treatment dosage with those that received no treatment at all. In our example, units in quadrant T_2 might receive a stronger dosage of treatment because they missed both the *R* and *M* cutoffs (Figure 1). However, the MRDD does not include these observations in the calculation of treatment effects because they are not located at F_M or F_R. A researcher may choose to redefine observations in T_2 as the treatment units, and those in the remaining three quadrants as the

comparison cases. However, this treatment contrast would involve comparing outcomes for units that missed both the *R* and *M* cutoffs (treatment cases) with those that missed only one cutoff (comparison cases). Again, this may not be the desired treatment contrast because treatment units are compared to those that have received at least some treatment because they missed either the *M* or *R* cutoffs. This predicament highlights one of the main design disadvantages of the MRDD, and one that researchers should be aware of as they interpret their treatment effects.

As with all Monte Carlo simulations, our study is limited by the fact that we cannot address every scenario that researchers are likely to encounter in analyzing MRDDs. First, we only examined the performance of the four approaches in the context of the sharp MRD design, where we assumed no instances of treatment misallocation. However, in many applications of MRDD, treatment crossover and no-show are likely to occur. Traditionally, fuzziness around the RD cutoff is addressed by using the assignment mechanism as an instrument for treatment receipt. This method extends to the MRD design for the univariate and centering approaches, where the treatment assignment mechanism again serves as an instrument for treatment receipt to estimate the local average treatment effect among compliers at the treatment frontier. Because of dimensionality issues, using an instrumental variable to address non-compliance for the frontier approach seems to be more challenging. A second limitation is that our study focuses on estimating treatment effects for MRDDs with only two assignment variables. Although the univariate, frontier, and centering approaches generalize well to MRDDs with more than two assignment variables, further work is needed for examining whether there are special analytic issues and requirements for analyzing MRDDs with more

complex assignment mechanisms (e.g., with exemption rules like those in No Child Left Behind). Finally, our simulation does not examine how variation in the correlation of the assignment variables would affect the performance of the four proposed methods. In particular, one might be concerned about the efficiency of treatment effect estimates generated by the univariate approach when the correlation between the two assignment variables are high, as would be the case in many education MRDDs where reading and math test scores are used as assignment variables.

7. CONCLUSION

This study defines causal estimands estimated by a multivariate regression-discontinuity design, and assesses the contexts and conditions under which four proposed approaches yield unbiased effect estimates. Results showed that the frontier, centering, univariate, and IV approaches succeed in producing unbiased treatment effects when their design and analytic assumptions are met. The IV approach, however, had stringent analytic requirements for yielding unbiased treatment effects, and produced inefficient results in our simulation study. The centering approach also yielded mixed results, but with small biases that are due to its increased sensitivity to bandwidth and functional form choices in local polynomial regressions. Finally, we found that the estimated causal quantity for the overall effect τ_{MRD} depended strongly on the scaling of the assignment variables, raising questions about the interpretability of τ_{MRD} when assignment variables are measured on different scales. Estimates for the frontier-specific effects, however, are not sensitive to the scaling and distribution properties of the assignment variables and provide lower and upper boundaries for the overall treatment effect.

REFERENCES

Aiken, L. S., West, S. G., Schwalm, D. E., Carroll, J., & Hsuing, S. (1998). Comparison of a randomized and two quasi-experimental designs in a single outcome evaluation: Efficacy of a university-level remedial writing program. *Evaluation Review, 22*, 207-244.

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics, 144*, 533-576.

Berk, R., Barnes, G., Ahlman, L., & Kurtz (in press). When second best is good enough: A comparison between a true experiment and a regressiondiscontinuity quasi-experiment. *Journal of Experimental Criminology*.

Berk, R. A., & de Leeuw, J. (1999). An evaluation of california's inmate classification system using a generalized regression discontinuity design. *Journal of the American Statistical Association, 94*(448), 1045-1052.

Berk, R. A., & Rauma, D. (1983). Capitalizing on nonrandom assignment to treatments: A regression-discontinuity evaluation of a crime-control program. *Journal of the American Statistical Association, 78*(381), 21-27.

Black, D., Galdo, J., & Smith, J. A. (2007). Evaluating the regression discontinuity design using experimental data. Retrieved from economics.uwo.ca/newsletter/misc/2009/smith_mar25.pdf

Buddelmeyer, H., & Skoufias, E. (2004). An evaluation of the performance of regression discontinuity design on PROGRESA. *World Bank Policy Research Working Paper No. 3386; IZA Discussion Paper No. 827*. Retrieved from http://ssrn.com/abstract=434600

Card, D., Chetty, R., & Weber, A. (2007). Cash-on-hand and competing models of intertemporal behavior: New evidence from the labor market. *Quarterly Journal of Economics, 122*(4), 1511-1560.

Cohen, J. L. (2006). Causes and consequences of special education placement: Evidence from Chicago Public Schools. Massachusetts Institute of Technology.

Cook, T. D., Wong, V. C., Steiner, P. M., Taylor, J., Gandhi, A., Kendziora, K., et al. (2009). *Impacts of School Improvement Status on Students with Disabilities: Feasibility report*. Washington, DC: American Institutes for Research.

Finkelstein, M., Levin, B., & Robbins, H. (1996a). Clinical and prophylactic trials with assured new treatment for those at greater risk: I. A design proposal. *Journal of Public Health, 86*(5), 691-695.

Finkelstein, M., Levin, B., & Robbins, H. (1996b). Clinical and prophylactic trials with assured new treatment for those at greater risk: II. Examples. *Journal of Public Health, 86*(5), 696-705.

Gamse, B., St. Pierre, B., Kemple, J., & Bloom, H. S. (2008). Reading First Impact Study Retrieved March 29, 2007, from http://ies.ed.gov/ncee/projects/evaluation/literacy_readingfirst.asp

Gill, B., Lockwood, J. R., Martorell, F., Setodji, C. M., & Booker, K. (2007). State and local implementation of the No Child Left Behind Act. U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

Goldberger, A. S. (2008). Selection bias in evaluating treatment effects: Some formal illustrations. In *Modelling and Evaluating Treatment Effects in Economics*, eds. T. Fomby, R. C. Hill, D. L. Millimet, J. A. Smith, E. J. Vytlacil, Amsterdam: JAI Press, pp. 1-31 .

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology, 41*(6), 872-884.

Gormley, W. T., & Phillips, D. (2005). The effects of universal pre-K in Oklahoma: Research highlights and policy implications. *The Policy Studies Journal 33*, 65-81.

Hahn, J., Todd, P., & van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica, 69*(1), 201-209.

Holland, P. W. (1986). Statistics and causal inference. *Journal of American Statistical Association, 81*(396), 945-960.

Imbens, G. W., & Lemieux, T. (2007). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*(2), 615-635.

Imbens, G., & Kalyanaraman, K. (2010). Optimal bandwidth choice for the regression discontinuity estimator. *Unpublished manuscript*.

Jackson, R., McCoy, A., Pistorino, C., Wilkinson, A., Burghardt, J., Clark, M., et al. (2007). *National evaluation of Early Reading First: Final report*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Jacob, B., & Lefgren, L. (2004a). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics, LXXXVI*, 226-244.

Jacob, B., & Lefgren, L. (2004b). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources, 39*, 50-79.

Kane, T. J. (2003). A quasi-experimental estimate of the impact of financial aid on college-going. Working Paper 9703, National Bureau of Economic Research.

Lalive, R. (2008). Unemployment benefits, unemployment duration, and post-unemployment jobs: A regression discontinuity approach. *American Economic Review, 97*(2), 108-112.

Lalive, R., Van Ours, J. C., & Zweimüller, J. (2006). How changes in fiscal incentives affect the duration of unemployment. *Review of Economic Studies, 73*(4), 1009-1038.

Lee, D. S., & Lemieux, T. (2009). Regression discontinuity designs in economics. NBER Working Paper 14723.

Ludwig, J., & Miller, D. L. (2005). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *NBER Working Paper No. 11702*.

Matsudaira, J. (2008). Mandatory summer school and student achievement. *Journal of Econometrics, 142*(2), 829-850.

Papay, J. P., Murnane, R. J., & Willett, J. B. (in press). The consequences of high school exit examiniations for low-performing urban students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis*.

Robinson, J. P., & Reardon, S. F. (2009). *Multiple regression discontinuity design: Implementation issues and empirical examples from education*. Paper presented at the Society for Research on Educational Effectiveness.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*, 688-701.

Shadish, W. S., Galindo, R., Wong, V. C., Steiner, P. M., & Cook, T. D. (under review). A Randomized Experiment Comparing Random to Cutoff-Based Assignment. *Psychological Methods*.

Trochim, W. M. K. (1984). *Research design for program evaluation*. Beverly Hills, CA: Sage Publications.

Van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A gegression-discontinuity approach. *International Economic Review, 43*(4), 1249-1287.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management, 27*(1), 122-154.

APPENDIX A

Proof of decomposition. Let $g(r,m) = y_1(r,m) - y_0(r,m)$ be the difference in potential treatment and control outcomes and $f(r,m)$ the joint density of *R* and *M*. Then, using line integration along frontier *F*, with *ds* being an infinitesimal segment along the frontier *F*, the expected treatment effect given *F* can be decomposed into the weighted average of conditional expectations:

$$\begin{aligned}\tau_{MRD} = E[G \mid F] &= \frac{\int_F g \cdot f \, ds}{\int_F f \, ds} = \frac{\int_{F_R} g \cdot f \, ds + \int_{F_M} g \cdot f \, ds}{\int_{F_R} f \, ds + \int_{F_M} f \, ds} \\ &= \frac{\int_{F_R} f \, ds}{\int_{F_R} f \, ds + \int_{F_M} f \, ds} \cdot \frac{\int_{F_R} g \cdot f \, ds}{\int_{F_R} f \, ds} + \frac{\int_{F_M} f \, ds}{\int_{F_R} f \, ds + \int_{F_M} f \, ds} \cdot \frac{\int_{F_M} g \cdot f \, ds}{\int_{F_M} f \, ds} \\ &= w_R \cdot E[G \mid F_R] + w_M \cdot E[G \mid F_M].\end{aligned}$$

Since either *R* or *M* but never both assignment variables change simultaneously along the cutoff frontier, we can rewrite the line integrals defining weights and conditional expectations in terms of regular integrals along each of the assignment variables. We rewrite the treatment weights as

$$w_R = \frac{\int_{m \geq m_c} f(r = r_c, m) dm}{\int_{m \geq m_c} f(r = r_c, m) dm + \int_{r \geq r_c} f(r, m = m_c) dr}, \tag{A1}$$

$$w_M = \frac{\int_{r \geq r_c} f(r, m = m_c) dr}{\int_{m \geq m_c} f(r = r_c, m) dm + \int_{r \geq r_c} f(r, m = m_c) dr},$$

and the conditional expectations as

$$E[G \mid F_R] = \frac{\int_{m \geq m_c} g(r,m) f(r = r_c, m) dm}{\int_{m \geq m_c} f(r = r_c, m) dm}, \text{ and} \tag{A2}$$

$$E[G \mid F_M] = \frac{\int_{r \geq r_c} g(r,m) f(r, m = m_c) dr}{\int_{r \geq r_c} f(r, m = m_c) dr}.$$

To simplify estimation of treatment weights and conditional expectations, it is useful to replace the bivariate density *f* by the product of the univariate conditional and marginal densities, $f = f_{R|M} \cdot f_M$ and $f = f_{M|R} \cdot f_R$, such that the treatment weights are given by

$$w_R = \frac{\int_{m \geq m_c} f_{M|R}(m \mid r = r_c) dm \cdot f_R(r = r_c)}{\int_{m \geq m_c} f_{M|R}(m \mid r = r_c) dm \cdot f_R(r = r_c) + \int_{r \geq r_c} f_{R|M}(r \mid m = m_c) dr \cdot f_M(m = m_c)} \text{ and} \tag{A3}$$

$$w_M = \frac{\int_{r \geq r_c} f_{R|M}(r \mid m = m_c) dr \cdot f_M(m = m_c)}{\int_{m \geq m_c} f_{M|R}(m \mid r = r_c) dm \cdot f_R(r = r_c) + \int_{r \geq r_c} f_{R|M}(r \mid m = m_c) dr \cdot f_M(m = m_c)},$$

and conditional expectations by

$$E[G \mid F_R] = \frac{\int_{m \geq m_c} g(r = r_c, m) \cdot f_{M|R}(m \mid r = r_c) dm}{\int_{m \geq m_c} f_{M|R}(m \mid r = r_c) dm} \text{ and} \tag{A4}$$

$$E[G \mid F_M] = \frac{\int_{r \geq r_c} g(r, m = m_c) \cdot f_{R|M}(r \mid m = m_c) dr}{\int_{r \geq r_c} f_{R|M}(r \mid m = m_c) dr}.$$

Figure 1. MRDD with two assignment variables *R* and *M*

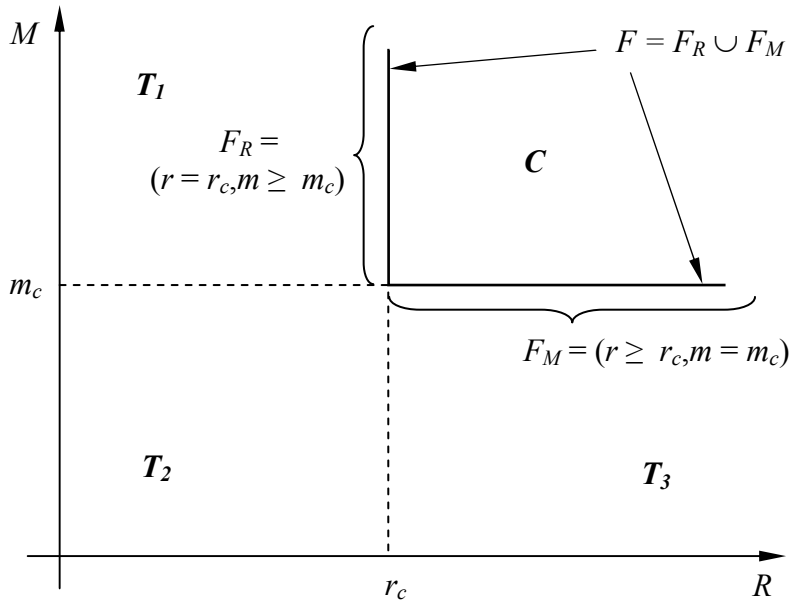

Figure 2. Response surface for potential treatment outcome $Y(1)$ without discontinuity (left panel) and with dicontiunities (right panel).

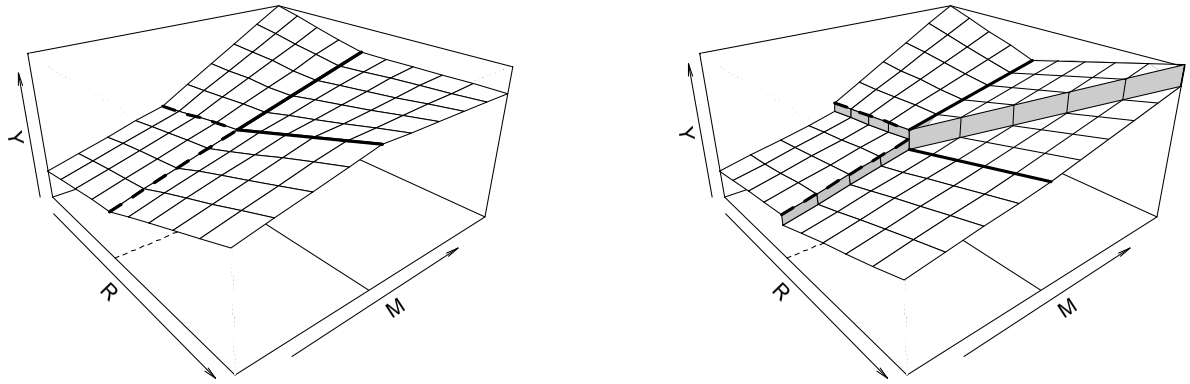

Figure 3. Scale-dependency of the joint density and weights at the cutoff frontier: By rescaling R such that $\text{var}(R_s) = 4\sigma_R^2$ the ratio of weights—represented by the ratio of the two areas along the frontier—changes.

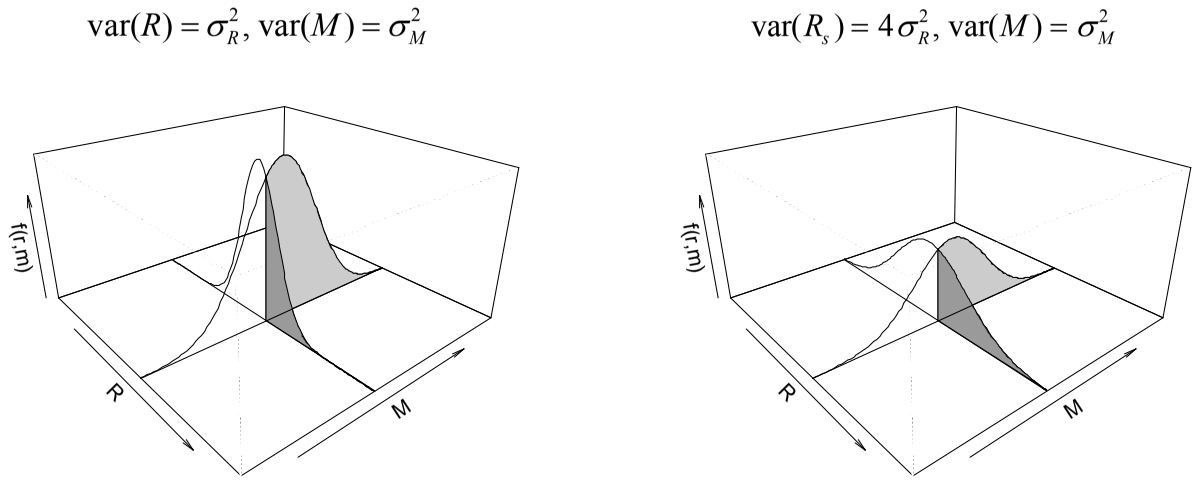

Figure 4. Response surfaces of observed outcomes for simulations: Constant treatment effects model (model 1), heterogeneous treatment effects model (model 2), and heterogeneous effects but with discontinuities in the potential treatment outcome model (model 3).

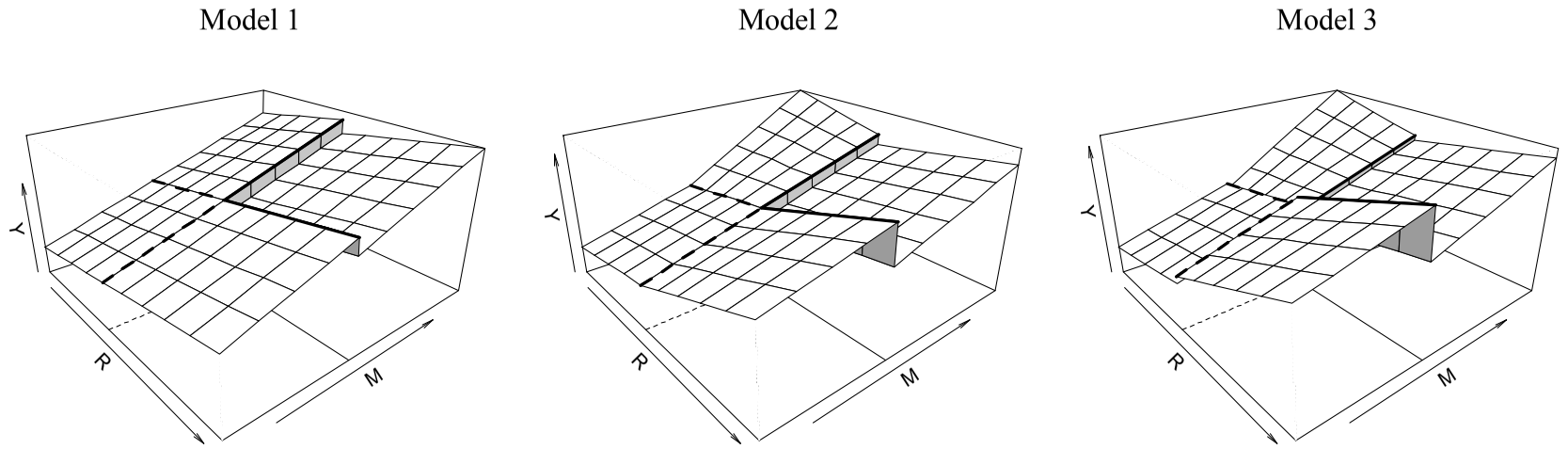

Table 1: MRD estimates for model 1 (constant treatment effects).

Assignment variables on	*same scale & same distribution*					*same scale & different distributions*					*different scales & distributions*				
Weights & effects	w_M	w_R	τ_M	τ_R	τ_{MRD}	w_M	w_R	τ_M	τ_R	τ_{MRD}	w_M	w_R	τ_M	τ_R	τ_{MRD}
True effect: raw AVs	73.0	27.0	4.00	4.00	4.00	51.5	48.5	4.00	4.00	4.00	99.6	0.4	4.00	4.00	4.00
True effect: standard.	73.0	27.0	4.00	4.00	4.00	80.9	19.1	4.00	4.00	4.00	73.0	27.0	4.00	4.00	4.00
Frontier approach:	72.8	27.2	3.99	4.01	4.00	51.3	48.7	3.99	4.02	4.01	99.6*	0.4*	4.00	4.01	4.00
full model	(3.1)	(3.1)	(0.1)	(0.19)	(0.10)	(4.5)	(4.5)	(0.1)	(0.23)	(0.13)	(0.1)	(0.1)	(0.1)	(0.20)	(0.10)
Frontier approach:	72.8	27.2	4.00	4.00	4.00	51.3	48.7	4.00	4.00	4.00	99.6*	0.4*	4.00	4.00	4.00
const. treat. effect	(3.1)	(3.1)	(0.09)	(0.09)	(0.09)	(4.5)	(4.5)	(0.09)	(0.09)	(0.09)	(0.1)	(0.1)	(0.09)	(0.09)	(0.09)
Centering approach:					3.98					3.86*					4.02
raw AVs					(0.47)					(0.90)					(0.28)
Centering approach:					3.94*					4.00					3.97*
standardized AVs					(0.35)					(0.50)					(0.36)
Univariate approach			3.98	3.99				3.99	4.13				3.99	4.01	
			(0.50)	(1.02)				(0.43)	(1.68)				(0.38)	(0.42)	
IV approach			4.06	4.31				4.00	4.32				3.98	4.09	
			(0.81)	(3.65)				(0.57)	(5.76)				(0.63)	(1.03)	

Notes. Weights w_M and w_R are given in percent ($w_M + w_R = 100$). Standard errors (estimated from the simulation) are in parenthesis.

* indicates statistical difference ($p < .05$) between estimated and theoretical treatment effects (based on simulation standard errors).

Table 2: MRD estimates for model 2 (heterogeneous treatment effects).

Assignment variables on	*same scale & same distribution*					*same scale & different distributions*					*different scales & distributions*				
Weights & effects	w_M	w_R	τ_M	τ_R	τ_{MRD}	w_M	w_R	τ_M	τ_R	τ_{MRD}	w_M	w_R	τ_M	τ_R	τ_{MRD}
True effect: raw AVs	73.0	27.0	9.74	3.70	8.11	51.5	48.5	7.59	3.36	5.54	99.6	0.4	9.74	4.00	9.72
True effect: standard.	73.0	27.0	9.74	3.70	8.11	80.9	19.1	7.59	3.36	6.79	73.0	27.0	9.74	4.00	8.19
Frontier approach:	72.5*	27.5*	9.77*	3.71	8.10	51.4	48.6	7.60	3.36	5.55	99.6	0.4	9.76	4.00	9.74
full model	(3.1)	(3.1)	(0.28)	(0.20)	(0.31)	(4.5)	(4.5)	(0.17)	(0.24)	(0.26)	(0.1)	(0.1)	(0.27)	(0.19)	(0.27)
Frontier approach:	72.5*	27.5*	9.80*	9.80*	9.80*	51.4	48.6	7.57*	7.57*	7.57*	99.6	0.4	9.43*	9.43*	9.43*
const. treat. effect	(3.1)	(3.1)	(0.13)	(0.13)	(0.13)	(4.5)	(4.5)	(0.11)	(0.11)	(0.11)	(0.1)	(0.1)	(0.12)	(0.12)	(0.12)
Centering approach:					8.10					5.45*					9.75
raw AVs					(0.64)					(0.84)					(0.37)
Centering approach:					8.07					6.80					8.12*
standardized AVs					(0.52)					(0.48)					(0.52)
Univariate approach			9.78	3.70				7.63	3.35				9.78	3.99	
			(0.73)	(0.97)				(0.54)	(1.51)				(0.55)	(0.40)	
IV approach			9.87*	3.80				7.61	3.35				9.82	4.00	
			(1.24)	(3.71)				(0.74)	(5.61)				(0.92)	(0.97)	

Notes. Weights w_M and w_R are given in percent ($w_M + w_R = 100$). Standard errors (estimated from the simulation) are in parenthesis.

* indicates statistical difference ($p < .05$) between estimated and theoretical treatment effects (based on simulation standard errors).

Table 3: MRD estimates for model 3 (heterogeneous effects but with discontinuities in the potential treatment outcome).

Assignment variables on	*same scale & same distribution*					*same scale & different distributions*					*different scales & distributions*				
Weights & effects	w_M	w_R	τ_M	τ_R	τ_{MRD}	w_M	w_R	τ_M	τ_R	τ_{MRD}	w_M	w_R	τ_M	τ_R	τ_{MRD}
True effect: raw AVs	73.0	27.0	8.61	1.70	6.74	51.5	48.5	7.63	1.36	4.59	99.6	0.4	8.61	2.00	8.59
True effect: standard.	73.0	27.0	8.61	1.70	6.74	80.9	19.1	7.63	1.36	6.44	73.0	27.0	8.61	2.00	6.82
Frontier approach:	72.9	27.1	8.62	1.70	6.75	51.2	48.8	7.64	1.36	4.59	99.6	0.4	8.62	1.99	8.59
full model	(3.3)	(3.3)	(0.16)	(0.19)	(0.28)	(5.0)	(5.0)	(0.11)	(0.23)	(0.35)	(0.1)	(0.1)	(0.16)	(0.20)	(0.16)
Frontier approach:	72.9	27.1	7.77*	7.77*	7.77*	51.2	48.8	6.83*	6.83*	6.83*	99.6	0.4	7.42*	7.42*	7.42*
const. treat. effect	(3.3)	(3.3)	(0.11)	(0.11)	(0.11)	(5.0)	(5.0)	(0.10)	(0.10)	(0.10)	(0.1)	(0.1)	(0.11)	(0.11)	(0.11)
Centering approach:					6.75					4.58					8.59
raw AVs					(0.52)					(0.88)					(0.30)
Centering approach:					6.73					6.49*					6.77*
standardized AVs					(0.41)					(0.47)					(0.46)
Univariate approach			8.58	1.72				7.63	1.40				8.59	2.00	
			(0.59)	(0.90)				(0.47)	(1.61)				(0.47)	(0.43)	
IV approach			9.42*	3.80*				7.96*	2.70*				9.33*	3.50*	
			(1.22)	(3.11)				(0.84)	(5.10)				(0.89)	(1.72)	

Notes. Weights w_M and w_R are given in percent ($w_M + w_R = 100$). Standard errors (estimated from the simulation) are in parenthesis.

* indicates statistical difference ($p < .05$) between estimated and theoretical treatment effects (based on simulation standard errors).