IDENTIFYING THE EFFECT OF CHANGING THE POLICY THRESHOLD IN REGRESSION DISCONTINUITY MODELS

Yingying Dong and Arthur Lewbel*

Abstract—Regression discontinuity models are commonly used to nonparametrically identify and estimate a local average treatment effect (LATE). We show that the derivative of the treatment effect with respect to the running variable at the cutoff, referred to as the treatment effect derivative (TED), is nonparametrically identified, easily estimated, and has implications for testing external validity and extrapolating the estimated LATE away from the cutoff. Given a local policy invariance assumption, we further show this TED equals the change in the treatment effect that would result from a marginal change in the threshold, which we call the marginal threshold treatment effect (MTTE). We apply these results to Goodman (2008), who estimates the effect of a scholarship program on college choice. MTTE in this case identifies how this treatment effect would change if the test score threshold to qualify for a scholarship were changed, even though no such change in threshold is actually observed.

I. Introduction

CONSIDER a standard regression discontinuity (RD) model, where T is a binary treatment indicator, X is a so-called running or forcing variable, c is the threshold for X at which the probability of treatment changes discretely, and Y is some observed outcome that may be affected both by treatment and smoothly by X. The goal in these models is to estimate the effect of treatment T on the outcome Y, and the main result in this literature is that under weak conditions, a local average treatment effect (LATE) can be nonparametrically identified and estimated at the point where $X = c$ (see, Hahn, Todd, & Van der Klaauw, 2001).

RD models identify a treatment effect locally at one point. Despite its strong internal validity, researchers often question the external validity of an estimated LATE. Therefore, it is useful to know what the treatment effect would be at points other than the cutoff c. For example, if the effect were very different at only slightly different values of X, then the external validity of the estimate should be a concern. But if the identified LATE is locally constant or nearly so, then it is more likely to have external validity. A related but separate issue is how the RD LATE would change if the RD threshold were marginally changed, since many policy questions center precisely on eligibility threshold changes.

In this paper, we show that one can nonparametrically identify a derivative of the RD LATE under weak conditions. The derivative can be used to explore the external validity of RD LATE in the neighborhood of the RD threshold, extrapolate the RD LATE away from the cutoff, and investigate how the RD LATE would change if the RD threshold were marginally changed.

To allow the treatment effect to vary with X, let $\pi(x)$ denote the average treatment effect (for compliers) when the running variable X equals the value x, and let $\pi'(x) = \partial\pi(x)/\partial x$ when this derivative exists. Note that the function $\pi(x)$ is defined holding the threshold fixed at c. RD estimation identifies $\pi(c)$, the LATE at $X = c$. We define TED (treatment effect derivative) as $\pi'(c)$. Intuitively, one can think of the derivative $\pi'(c)$ as the coefficient of the interaction term between the treatment T and $X-c$ in a (local) linear regression of Y on a constant, T, $X - c$, and $(X - c)T$.

We show that TED for both sharp designs and fuzzy designs can be nonparametrically identified and is easily estimated. The smoothness conditions needed for identifying TED are slightly stronger than those needed to identify $\pi(c)$. However, inference for standard methods used to actually estimate $\pi(c)$, such as kernel or local polynomial-based estimators, or standard parametric functional forms, requires precisely the same additional smoothness assumptions that are needed to identify and estimate TED. As a result, all empirical applications of RD methods that we know of already make the assumptions needed to identify TED.

One use of TED is to test for locally constant treatment effects, since $\pi'(c) = 0$ is a necessary condition for not having the treatment effect change with the running variable. More generally, if $\pi'(c)$ is large in magnitude, then a small change in the running variable is associated with a large change in treatment effect, which would then call into question the external validity or generality of the estimated LATE. The sign of TED is also informative, since it tells whether the treatment effect is likely to be larger or smaller for individuals (units) with a value of X that is slightly larger or smaller than c.

In many RD applications, the function $E(Y \mid X = x)$ is parameterized; for example, this function is assumed to be a polynomial in chapter 6 of Angrist and Pischke (2008). When this expectation is parameterized, the function $\pi(x)$ can be obtained for all values of x. For example, suppose we have a sharp design model where this expectation is assumed to be quadratic, so $Y = a + Xb + X^2d + T\tilde{a} + XT\tilde{b} + X^2T\tilde{d} + e$ and $E(e \mid X = x) = 0$ for all x in some interval. Then for this model, $\pi(x) = \tilde{a} + x\tilde{b} + x^2\tilde{d}$, so $\pi'(x) = \tilde{b} + 2x\tilde{d}$. So in this case, TED is given by $\tilde{b} + 2x\tilde{d}$ and is thereby identified for all x in an interval. But is this identification due to the assumed functional form, or can TED be nonparametrically identified?

In parametric models $\pi(x)$ is identified both at $x = c$ and for values $x \neq c$, permitting identification of $\pi'(c)$ only because the functional form allows us to evaluate counterfactual objects like $E(Y \mid T = 1, X = x < c)$, even though in the

Received for publication November 22, 2013. Revision accepted for publication September 3, 2014. Editor: Mark W. Watson.

* Dong: University of California Irvine; Lewbel: Boston College.

We thank Josh Angrist and some anonymous referees for helpful comments, and thank Joshua Goodman and Justin Gallagher for kindly providing data and related help. All errors are our own. Early versions of this paper were circulated under the title "Regression Discontinuity Marginal Threshold Treatment Effects."

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/REST_a_00510.

doi:10.1162/REST_a_00510

data with a sharp design, we could never see any observations having both $T = 1$ and $x < c$. One might think that nothing regarding changes in treatment effects can be identified nonparametrically, because we observe treatment change only at $x = c$ itself.

However, what we show is that, analogous to the way $\pi(c)$ in sharp designs is nonparametrically identified by differencing the left and right limits of $E(Y \mid X = x)$ as $x \to c$, one can similarly nonparametrically identify TED, $\pi'(x)$ at the point $x = c$, by differencing the left and right derivatives of $E(Y \mid X = x)$ with respect to x as $x \to c$. We also show that TED can be identified and easily estimated in fuzzy designs as well.

A separate issue from identification of TED is consideration of the threshold c. In many applications, the threshold is itself a policy variable of interest. Knowing how the direction and magnitude of the treatment effect would change if the threshold were changed can be important in practice. Policy debates often center precisely on these types of questions (some examples are given below).

In order to consider the impacts of changing the policy threshold, let C denote a possible threshold value. The observed threshold value is $C = c$. Further let $\tau(C)$ denote the LATE that would be identified by a standard RD model if the threshold equaled C, so $\tau(C)$ is the average treatment effect for compliers at the cutoff $X = C$. Note that this differs from measuring treatment effect heterogeneity in the running variable as TED does; instead we are now considering how the LATE would change if the RD threshold changed. Let $\tau'(C) = \partial\tau(C)/\partial C$. We define MTTE, the marginal threshold treatment effect, to be this derivative at the actual threshold c, so MTTE is $\tau'(c)$. We show that if a certain local policy invariance condition is satisfied, then MTTE equals TED, that is, $\tau'(c) = \pi'(c)$, so when local policy invariance holds, MTTE is identified and easily estimated.

To see the link between TED and MTTE, let $S(x, c)$ be an average treatment effect for individuals having running variable $X = x$ when the policy threshold for assigning treatment is $C = c$, so the first argument of $S(x, c)$ allows for treatment effect heterogeneity in the running variable, and the second argument allows the treatment effect to depend on the RD threshold. The difference between the functions τ and D is that whatever the true threshold c is, $\tau(c) = S(c, c)$, while $\pi(x) = S(x, c)$. It follows that TED is given by $\pi'(c) = \frac{\partial S(X,c)}{\partial X}\,|_{X=c}$, while MTTE is $\tau'(c) = \frac{\partial S(C,C)}{\partial C}\,|_{C=c} = \frac{\partial S(X,C)}{\partial X}\,|_{X=c,C=c} + \frac{\partial S(X,C)}{\partial C}\,|_{X=c,C=c}$. The local policy invariance assumption is that $\frac{\partial S(X,C)}{\partial C}\,|_{X=c,C=c} = 0$, under which MTTE equals TED.

Local policy invariance assumes that the function that describes how the RD LATE varies with the running variable does not itself change when the policy threshold changes infinitesimally. It is essentially a ceteris paribus assumption of the type commonly employed in partial equilibrium analyses. Analogous ceteris paribus assumptions are required to apply almost any reduced-form program evaluation calculations to a change in context or environment.

Abbring and Heckman (2007) define *policy invariance* as an "assumption that an agent's outcome only depends on the treatment assigned to the agent and not separately on the mechanism used to assign treatments. This excludes (strategic) interactions between agents and equilibrium effects of the policy." Example applications of policy invariance assumptions and marginal policy analyses include Heckman (2010) and Carneiro, Heckman, and Vytlacil (2010). What we require to identify MTTE is a limited version of policy invariance that applies only in response to an infinitesimal change in the assignment mechanism, a change in c, and which we therefore refer to as "local policy invariance."

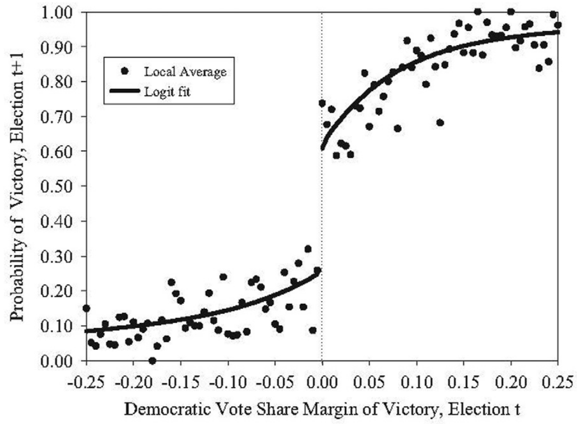
MTTE can be used to approximate the impact on treatment effects of a small discrete change in the threshold exactly the way that, for example, price elasticities are used to approximate the effects of small discrete changes in prices. The sign of MTTE tells whether the average effectiveness of treatment would increase or decrease if the threshold for treatment were marginally changed.

In discussing RD methods, Hahn, Todd, and Van der Klaauw (2001) note that "a limitation of the approach is that it only identifies treatment effects locally at the point at which the probability of receiving treatment changes discontinuously. . . . It would be of interest, for example, if the policy change being considered is a small change in the program rules, such as lowering or raising the threshold for program entry, in which case we would want to know the effect of treatment for the subpopulation affected by the change." In estimating the effect of Superfund-sponsored cleanups of hazardous waste sites on housing prices, Greenstone and Gallagher (2008) have noted, "It is important to highlight that the RD approach only provides estimates of the treatment effect at the regulatory discontinuity (i.e., HRS = 28.5). To extend the external validity of the RD estimates to the full 1982 HRS Sample, it is necessary to assume a homogeneous treatment effect in that sample." MTTE and TED discussed here address these issues by showing how the average treatment effect would change at slightly different values of the running variable other than the RD threshold or given a marginal change in the threshold.

To illustrate the usefulness of TED and MTTE, consider some empirical applications. One example is Lee (2008), who uses a sharp design RD to estimate a party's incumbency advantage in U.S. House of Representatives elections. The treatment T is having the incumbent be a Democrat,[1] X is the Democratic's winning margin (the difference between the Democratic Party's vote share and its strongest opponent share in election t), and Y is the probability of a Democrat winning in election $t + 1$. Figure 1, which is reproduced from figure 5a in Lee (2008), shows how this probability changes with the winning margin in election t. Notice in particular that the slope is steeper just to the right of the discontinuity than on the left. TED is essentially the difference in these

[1] Due to the largely two-party system, the strongest opponent is almost always a Republican, and the outcome for the Republican Party is therefore a mirror image of Democratic's outcome (Lee, 2008).

FIGURE 1.—PROBABILITY OF THE DEMOCRATIC PARTY WINNING ELECTION $t+1$ AGAINST ITS WINNING MARGIN IN ELECTION t



slopes, which in this application provides (for close past elections) a measure of how the incumbent's electoral advantage depends on its winning margin in the previous election. Figure 1 suggests that the incumbency advantage increases with the winning margin, meaning that the larger is the incumbent party's share in the previous election, the greater is their chance of winning in the next election.

Another example is Goodman (2008), who studies the effect of the Adams Scholarship program on college choices in Massachusetts. The scholarship program provides tuition wavers at in-state public colleges. The treatment T is Adams Scholarship eligibility, which is determined by whether a standardized test score exceeds a certain threshold c. The running variable X is the number of grade points one is from the eligibility threshold. One of the outcomes of interest Y is the probability of choosing a public college. TED, in this case (which we show later is negative and strongly significant), provides information on how students' responses to Adams Scholarship eligibility depend on their test scores, and hence on their outside opportunities. When policy invariance holds, MTTE then allows for evaluating how the average effects of the Adams Scholarship program would change if the qualifying threshold were marginally raised or lowered.

A third example is Greenstone and Gallagher (2008). The treatment T is an indicator for being eligible for Superfund-sponsored cleanups, which is largely determined by whether a hazardous ranking system (HRS) score assigned by the EPA in 1982 exceeds a certain threshold. The running variable X is a county's 1982 HRS score minus a threshold score, and the outcome Y is median housing prices in the surrounding area. TED in this case shows how the effect of Superfund cleanups on housing prices would change with the HRS score, measuring the hazard level of a waste site, and thereby provides information on the external validity of the estimated LATE at the current regulatory threshold. MTTE, when valid, would then tell us how the impact of Superfund cleanups would change if the threshold changed slightly.

A few other papers consider derivative conditions in RD analyses. Card et al. (2012) analyze a regression kink design model, where the treatment is a continuous but kinked function of the running variable. The kink, which is a discontinuity in a derivative, is used to identify a treatment effect. Dong (2014a) uses changes in the derivative or slope of the treatment probability at the threshold to identify standard fuzzy RD design treatment effects in applications where there is a kink instead of, or in addition to, a discontinuity in this probability at the threshold. Both Abdulkadiroglu, Angrist, and Pathak (2014) and Angrist and Rokkanen (2012) make use of TED, citing an earlier version of our paper. But perhaps the closest result to ours is a few paragraphs in a survey article by Dinardo and Lee (2011), in which they informally propose using a Taylor expansion at the threshold to identify a sharp design average treatment effect on the treated (ATT) parameter. In contrast, we use a similar expansion to estimate different objects (TED and MTTE), and we provide formal results for both fuzzy and sharp designs.

Also closely related is recent work on extrapolation away from the cutoff in RD models, such as Kirabo Jackson (2010), Angrist and Rokkanen (2012), and Maynard et al. (2013). These papers work by assuming the availability of additional variables or other information rather than through derivatives and local neighborhood assumptions. Although the methodology in these papers differs substantially from ours, their motivation is similar, further demonstrating the interest and value of estimating how treatment effects would change with the running variable x or the RD threshold c.

The rest of the paper is organized as follows. Sections II and III show nonparametric identification of TED in sharp and fuzzy RD designs, respectively. Section IV describes in detail the local policy invariance condition required to have MTTE equal TED. Section V discusses estimation of TED and MTTE. Section VI provides an empirical application based on Goodman (2008). Section VII concludes, and proofs are provided in the appendix. We also provide some additional theoretical results and a second empirical application based on Greenstone and Gallagher (2008) in a supplemental online appendix.

II. Sharp Design TED

This section discusses identification of TED in sharp design RD. As in Rubin (1974), let $Y(1)$ and $Y(0)$ denote the potential outcomes when one is treated or not treated, respectively, so the observed outcome is $Y = Y(1)T + Y(0)(1-T)$. Holding the threshold fixed at the value c, define $\pi(X) = E[Y(1) - Y(0) \mid X]$ and its derivative $\pi'(X) = \partial\pi(X)/\partial X$. For simplicity, we first provide assumptions and results without consideration of covariates other than the running variable X. We later discuss how additional covariates Z can be included.

Define $g(x) = E(Y \mid X = x)$. For small $\varepsilon > 0$, define the right and left limits of a given function h as $h_+(x) =$

$\lim_{\varepsilon \longrightarrow 0} h(x+\varepsilon)$ and $h_{-}(x) = \lim_{\varepsilon \longrightarrow 0} h(x-\varepsilon)$, respectively.

The main identification result in the literature for sharp design RD is that $\pi(c)$, the LATE for individuals having X equal to the observed cutoff c, is identified by[2]

$$\pi(c) = g_{+}(c) - g_{-}(c-\varepsilon) \tag{1}$$

based on, for $\varepsilon > 0$,

$$\begin{aligned} E(Y(0) \mid X = c) &= \lim_{\varepsilon \longrightarrow 0} E(Y(0) \mid X = c-\varepsilon) \\ &= \lim_{\varepsilon \longrightarrow 0} E(Y(0) \mid X = c-\varepsilon, T = 0) \\ &= \lim_{\varepsilon \longrightarrow 0} E(Y \mid X = c-\varepsilon) \end{aligned} \tag{2}$$

and similarly for $Y(1)$ using $X = c+\varepsilon$. Formally this result can be obtained given the following assumptions.

Assumption A1. For each unit (individual) i we observe Y_i, T_i, and X_i. $Y_i = (1-T_i)Y_i(0) + T_iY_i(1)$.

For ease of notation we drop the i subscript throughout.

Assumption A2 (sharp design). $T = I(X \geq c)$ for some known constant threshold c. The support of X includes a neighborhood of c. For $t = 0, 1$, $E(Y(t) \mid X = x)$ is continuously differentiable in x in a neighborhood of $x = c$.

Note that assumption A2 assumes not just continuity but differentiability, though continuity suffices to establish equation (1) by the derivation in equation (2). However, virtually all empirical implementations of RD models satisfy this stronger smoothness assumption. In particular, parametric models generally assume polynomials or other differentiable functions, while most nonparametric estimators, including local linear regressions, assume (for establishing asymptotic theory) at least continuous differentiability of $E(Y(1) \mid X = x)$ for $x \geq c$ and of $E(Y(0) \mid X = x)$ for $x \leq c$.[3] We do not know of any empirical application of RD methods where the usual continuity assumptions hold but where the additional differentiability of assumption A2 was not (either implicitly or explicitly for estimation) assumed to hold.[4]

To show identification of TED, we need one-sided derivatives. For small $\varepsilon > 0$, define the right and left derivatives of a function $h(x)$ at the point x as

$$h'_{+}(x) = \lim_{\varepsilon \longrightarrow 0} \frac{h(x+\varepsilon) - h(x)}{\varepsilon} \quad \text{and}$$
$$h'_{-}(x) = \lim_{\varepsilon \longrightarrow 0} \frac{h(x) - h(x-\varepsilon)}{\varepsilon}.$$

Theorem 1. *If assumptions A1 and A2 hold, then the treatment effect $\pi(c)$ holding the threshold fixed at c is identified by equation (1) and TED defined by $\pi'(c)$ is identified by*

$$\pi'(c) = g'_{+}(c) - g'_{-}(c). \tag{3}$$

Proofs are in the appendix. Estimation of TED based on theorem 1 will be discussed in more detail later, but for now, note that local polynomial regressions can be used to estimate the function $g(x)$ and its derivatives separately on either side of the threshold. Evaluating these regression derivatives in the limits as $x \rightarrow c$ provides consistent estimators of $g'_{+}(c)$ and $g'_{-}(c)$, and hence consistent estimators of $\pi'(c)$. In short, $\pi'(c)$ equals the difference between the left and right derivatives of $g(x)$ around $x = c$, just as the local treatment effect $\pi(c)$ equals the difference between the left and right limits of $g(x)$ around $x = c$.

As discussed in section I, TED provides a measure of the impact of a marginal change in the running variable x on the treatment effect, at $x = c$, holding c fixed. This estimate can be used for a variety of purposes. For example, the sign of TED tells whether an increase in the running variable is likely to be associated with an increase versus a decrease in the size of treatment effects. TED can be used to test for locally constant treatment effects, since having TED equal 0 is necessary for not having the treatment effect change with the running variable. More generally, if TED is nonnegligible in magnitude, then one might be concerned about the external validity of the estimated RD LATE, since a large TED suggests that even small changes in the running variable can be associated with large changes in treatment effects.

III. Fuzzy Design TED

We now extend our previous result to fuzzy designs. Let T continue to indicate whether an individual is treated, but now let $T^* = I(X \geq c)$, so T^* is a dummy indicating whether one is above or below the threshold. T would be the same as T^* for all individuals if the design were sharp. Define $f(x) = E(T \mid X = x)$, so $f(x)$ is the probability of being treated given $X = x$. Let the counterfactual outcomes $Y(t)$ for $T = t$ be defined as before. Analogous to $Y(t)$ in the sharp design, define the potential treatment status $T(t^*)$ as what an individual's treatment status would be if $T^* = t^*$.[5]

[2] See Hahn et al. (2001), or recent surveys Lee and Lemieux (2010), Imbens and Wooldridge (2009), and Imbens and Lemieux (2008).

[3] Local linear or higher-order local polynomial regressions are used to estimate RD models to mitigate boundary bias issues as discussed by, e.g., Porter (2003). The asymptotic theory for local linear or local polynomial estimation (see Fan & Gijbels, 1996) requires not just continuity but continuous differentiability. Generally twice differentiability or more smoothness is assumed. This differentiability is not required for consistent estimation of RD models, but is always assumed in practice to reduce bias and thereby increase estimation precision. Similarly, still greater smoothness might be useful for bias mitigation in derivative estimation but is not required for consistency of TED estimation.

[4] Spline estimators are not differentiable at the knots, but one could define a neighborhood around c to be smaller than the closest knot to c.

[5] Here we define the potential treatment notation $T_i(0)$ and $T_i(1)$ in terms of $T^* = I(X \geq c)$ only, so we implicitly assume that for any given individual i, $T_i(0)$ and $T_i(1)$ are constant in a neighborhood of $X_i = c$. This assumption be relaxed to allow $T_i(0)$ and $T_i(1)$ to futher be a function of X_i. See the discussion in Dong (2014a, 2014b). Relaxing this assumption would not materially change our derivations or conclusions.

So analogous to $Y = Y(1)T + Y(0)(1 - T)$, we have $T = T(1)T^* + T(0)(1 - T^*)$. Define compliers as individuals i having $T_i(0) < T_i(1)$ and defiers as those having $T_i(0) > T_i(1)$, always takers as those having $T_i(0) = T_i(1) = 1$, and never takers as those having $T_i(0) = T_i(1) = 0$.

Analogous to $\pi(x)$ in sharp designs, define

$$\pi_f(x) = E[Y(1) - Y(0) \mid X = x, T(0) < T(1)] \quad (4)$$

where the subscript f denotes fuzzy design. The only difference between the functions $\pi_f(x)$ and $\pi(x)$ is that the definition of $\pi_f(x)$ conditions explicitly on compliers, $T(0) < T(1)$. Both $\pi_f(x)$ and $\pi(x)$ implicitly condition on having the threshold fixed at c. The standard fuzzy design RD identification result as in Hahn et al. (2001) is that $\pi_f(x)$ at the point $x = c$ is identified by

$$\pi_f(c) = \frac{g_+(c) - g_-(c)}{f_+(c) - f_-(c)}. \quad (5)$$

Assumption A3 (fuzzy design). The support of X includes a neighborhood of a known constant threshold c. $T(0) \leq T(1)$. The conditional means $E(Y(t) \mid T(0) < T(1), X = x)$ and $E(Y(t) \mid T(0) = T(1) = t, X = x)$, as well as the probabilities $\Pr(T(0) < T(1) \mid X = x)$ and $\Pr(T(0) = T(1) = t \mid X = x)$ for $t = 0, 1$, are continuously differentiable in x in a neighborhood of $x = c$. $\Pr(T(0) < T(1) \mid X = x)$ is strictly positive at $x = c$.

Assumption A3 differs in two important ways from the standard assumptions used by, for example, Hahn et al. (2001), to obtain equation (5). The first difference is that this assumption assumes continuous differentiability instead of just continuity of some functions. Given differentiability, we can define $\pi'_f(x) = \partial \pi_f(x)/\partial x$, and TED for compliers is then $\pi'_f(c)$. As in assumption A2 and theorem 1, differentiability in assumption B2 is not needed to identify $\pi_f(c)$, but is always assumed in practice to do inference on associated estimators, and we now use it to identify the fuzzy design TED $\pi'_f(c)$.

The second key difference between assumption A3 and standard assumptions is that A3 does not assume that the treatment effect $Y(1) - Y(0)$ is independent of treatment T conditional on X near c, nor does it assume that the treatment effect and potential treatment status $(Y(1) - Y(0), T(x))$ are jointly independent of X near c (see theorem 2 and assumption A3 of theorem 3 in Hahn et al., 2001).

In place of independence, assumption A3 assumes smoothness of conditional means of potential outcomes for each type of individual and smoothness of probabilities of selection into each type. Dong (2014b) shows that these alternative assumptions (using just continuity, not differentiability) suffice to obtain equation (5). Dong (2014b) also provides a weak behavioral assumption that is sufficient to make these alternative smoothness conditions hold. Intuitively, the mean outcome right below or above the cutoff is a weighted average of the mean outcomes for each type of individual, weighted by the probabilities of each type. When the conditional means for each type and the related probabilities are all smooth at the cutoff, the mean outcome difference at the cutoff then just equals the mean change in outcomes for compliers.

We make these alternative smoothness assumptions because we want to allow the possibility that X may be correlated with $(Y(1) - Y(0), T(x))$ even when X is near c. In particular, if $(Y(1) - Y(0), T(x))$ were independent of X near c, then that would imply that TED is 0, a restriction we do not want or need to impose. Given our assumptions, the following theorem shows identification of TED:

Theorem 2. *If assumptions A1 and B2 hold, then the treatment effect $\pi_f(c)$ is identified by equation (5) and the fuzzy design RD TED $\pi'_f(c)$ is identified by*

$$\pi'_f(c) = \frac{g'_+(c) - g'_-(c) - \left[f'_+(c) - f'_-(c)\right]\pi_f(c)}{f_+(c) - f_-(c)}. \quad (6)$$

The same uses for TED that we discussed in sharp designs carry over to this fuzzy design TED. For example, a necessary condition to have the fuzzy design treatment effect $\pi_f(c)$ be locally constant is $\pi'_f(c) = 0$, so, for example, a finding that the estimated TED was significantly different from 0 would allow us to reject the hypothesis of a locally constant treatment effect.

Let $p(c)$ denote the fraction of the population at $X = c$ who are compliers. Applying theorem 1 to the treatment equation, treating T as the outcome and T^* as the treatment, shows that $p(c) = f_+(c) - f_-(c)$ and $p'(c) = f'_+(c) - f'_-(c)$. We can therefore test if the compliance rate is locally constant by testing if $p'(c)$ is 0. We can also then rewrite equation (6) as

$$\pi'_f(c) = \frac{g'_+(c) - g'_-(c)}{p(c)} - \frac{p'(c)\,\pi_f(c)}{p(c)}. \quad (7)$$

Equation (7) provides a way of interpreting equation (6). The first term on the right side of equation (7) is what the fuzzy design TED equals if the compliance rate is locally constant (equivalently, if the compliance rate was held fixed). In sharp designs, the probability of compliance is constant at 1, so $p(c) = 1$ and $p'(c) = 0$, and equation (7) then reduces to the sharp design TED in equation (3). The second term in equation (7) is proportional to $p'(c)$ and accounts for the effect on TED of a change in the probability of compliance.

IV. Sharp Design MTTE

We now consider how the treatment effect would change if the threshold changed. In general, the distributions of $Y(0)$ and $Y(1)$ can depend on both the running variable and the RD threshold. For example, a person's outcome if treated $Y(1)$ may depend on who else was treated, and thereby depend on the value of the threshold, in addition to directly depending on the value of that person's running variable. We let $\tau(C)$ denote the LATE that would be identified by RD methods if the threshold equaled C. The observed threshold is c, so the actual RD LATE we can identify is $\tau(c)$. The goal of this section is to identify the marginal threshold treatment effect or the function $\tau'(C)$ evaluated at $C = c$.

As in section I, define the sharp design function $S(X, C) = E(Y(1) - Y(0) \mid X, C)$, so $S(X, C)$ is the average treatment effect for hypothetical individuals having a running variable equal to X, living in a world where treatment is assigned based on whether X exceeds C. This $S(X, C)$, while well defined as a function, is only a hypothetical treatment effect when $X \neq C$ or $C \neq c$. Assuming S is differentiable, denote the derivatives of $S(X, C)$ by

$$S_X(X, C) = \frac{\partial S(X, C)}{\partial X}, \; S_C(X, C) = \frac{\partial S(X, C)}{\partial C}.$$

It follows from these definitions that TED $\pi'(c)$ and MTTE $\tau'(c)$ are given by

$$\pi'(c) = S_X(c, c) \quad \text{and} \quad \tau'(c) = \pi'(c) + S_C(c, c). \qquad (8)$$

We define sharp design local policy invariance to be the condition that $S_C(c, c) = 0$. Local policy invariance means that the change in the LATE function $\pi(x)$ at $x = c$, relative to a change ε in the true threshold c, shrinks to 0 as $\varepsilon \to 0$. A sufficient but stronger than necessary condition for this to hold would be if the LATE function $\pi(x)$, at points x in an arbitrarily small neighborhood of $x = c$, did not change in response to an infinitesimal change in the true threshold c. Note that local policy invariance does not imply, and is not implied by, locally constant treatment effects. In fact, local policy invariance places no restriction on the shape of the function $\pi(x)$. It only restricts how this function could change if the threshold changed marginally.

The following is an immediate implication of theorem 1 and equation (8):

Corollary 1. *Let assumptions A1 and A2 and the local policy invariance condition $S_C(c, c) = 0$ hold. Then MTTE is nonparametrically identified by $\tau'(c) = \pi'(c)$.*

Local policy invariance makes TED equal MTTE only at the point $C = c$, so in general we may have $\tau'(C) \neq \pi'(C)$ for $C \neq c$. Given MTTE, we can use the mean value theorem to obtain an approximate estimate of the effect of a small, discrete change in the threshold. Specifically, an estimate of what the treatment effect $\tau(c_{new})$ would be if the threshold were changed a small amount from c to c_{new} is given by $\tau(c_{new}) \approx \tau(c) + (c_{new} - c)\,\tau'(c)$. Suppose local policy invariance does not hold exactly, meaning that $S_C(c, c)$ is not precisely 0. If $S_C(c, c)$ is negligibly small (meaning small in magnitude relative to the approximation error in the $\tau(c_{new})$ application of the mean value theorem), then we can approximate the new threshold treatment effect $\tau(c_{new})$ by

$$\tau(c_{new}) \approx \pi(c) + (c_{new} - c)\,\pi'(c). \qquad (9)$$

Similarly, given bounds on $S_C(c, c)$, we could immediately construct corresponding bounds on $\tau(c_{new})$.

The plausibility of local policy invariance depends on context. To see why local policy invariance might not hold, consider the original Thistlethwaite and Campbell (1960) RD model, where T is receipt of a National Merit Award, X is the test score on the award-qualifying exam, c is the threshold grade required to qualify for the award, and Y is receipt of college scholarships later or other academic outcomes. Consider a group of compliers who have test scores x that equal a value c_{new} that is infinitesimally larger or smaller than c. A sufficient condition for local policy invariance is if the average treatment effect for students in this group, who already have $x = c_{new}$, would not change if the threshold test score used for determining treatment were changed from c to c_{new}. This assumption might be violated if, for example, changes in the number of award winners resulting from infinitesimal changes in the qualifying threshold c lead to increased or decreased competition for college scholarships, or lead to changes in the perceived prestige of a merit award.

For another example, consider an RD model like that of Jacob and Lefgren (2004). The treatment T is attending summer school or repeating a grade, which is determined by whether a standardized test score X falls below a threshold failing score c and the outcome is later academic performance. Local policy invariance here would hold if the effect of repeating a grade or attending summer school (as a function of X) did not change for compliers at a marginally lower or higher threshold. Such an assumption might not hold if a marginal change in the threshold failing score would affect the curriculum or quality of instruction, or if the resulting small compositional change in the population of students in the grade repeated or in the summer school affected learning through peer effects.

Conditioning on covariates could help in making local policy invariance hold. For example, in an application where treatment is attending summer school, if the only source of potential violation of local policy invariance is that a change in threshold could change summer school class sizes, then including class sizes as an additional covariate would suffice to satisfy a conditional local policy invariance restriction.

These examples illustrate cases where local policy invariance might be violated. In other contexts, local policy invariance is more likely to hold. We explain in detail the plausibility of local policy invariance (or at least plausibility of having $S_C(c, c)$ be negligibly small) later in our empirical applications, including the Adams Scholarship program in Massachusetts following Goodman (2008) and Superfund cleanups as analyzed by Greenstone and Gallagher (2008).

V. Fuzzy Design MTTE

The results of section IV extend immediately to fuzzy designs. Replace the function S and its derivatives there with the function S_f and its derivatives, defined as

$$S_f(X,C) = E(Y(1) - Y(0) \mid X, C, T(0) < T(1)),$$
$$S_{fX}(X,C) = \frac{\partial S_f(X,C)}{\partial X},\ S_{fC}(X,C) = \frac{\partial S_f(X,C)}{\partial C}.$$

We use the subscript f to denote fuzzy design. The only difference between functions S_f and S is that S_f also conditions on compliers—individuals with $T(0) < T(1)$. The standard fuzzy design RD LATE with a threshold C is $\tau_f(C) = S_f(C,C)$. The fuzzy design MTTE is then $\tau_f'(c)$. Analogous to equation (8), the fuzzy design TED and MTTE are given by

$$\pi_f'(c) = S_{fC}(c,c), \text{ and } \tau_f'(c) = \pi_f'(c) + S_{fC}(c,c). \tag{10}$$

Fuzzy design local policy invariance is the condition that $S_{fC}(c,c) = 0$. As in the sharp design case, local policy invariance does not place any restriction on how the treatment effect depends on the running variable, that is, it does not restrict the function $\pi_f(x)$.

Corollary 2. *Let assumptions A1, A3, and the local policy invariance condition $S_{fC}(c,c) = 0$ hold. Then the fuzzy design MTTE $\tau_f'(c)$ is nonparametrically identified by $\tau_f'(c) = \pi_f'(c)$.*

The immediate analog to equation (9) is that if $S_{fC}(c,c)$ is small, then $\tau_f(c_{new}) \approx \pi_f(c) + (c_{new} - c)\,\pi_f'(c)$, which allows us to assess what the RD treatment effect would equal if the threshold were changed slightly from c to c_{new}.

A difference between sharp and fuzzy designs for MTTE is that if the threshold changed, then in general, the set of people who are compliers would change. This is reflected in equations (6) and (7). As discussed previously, in fuzzy designs, the denominator $f_+(c) - f_-(c)$ represents the compliance rate $p(c)$. Therefore, given local policy invariance, $p'(c) = f_+'(c) - f_-'(c)$ equals the change in the compliance rate that would result if the threshold changed marginally. By applying the mean value theorem as before, one could approximate the compliance rate at a new threshold c_{new} by

$$p(c_{new}) \approx p(c) + (c_{new} - c)\,p'(c). \tag{11}$$

So even though the fraction of the population who would be compliers can change in unknown ways when the threshold changes, local policy invariance allows us to approximate the compliance rate and the treatment effect, $p(c_{new})$ and $\tau_f(c_{new})$, respectively, at the new threshold.

Local policy invariance along with equation (7) gives

$$\tau_f'(c) = \frac{g_+'(c) - g_-'(c)}{p(c)} - \frac{p'(c)\,\pi_f(c)}{p(c)}. \tag{12}$$

VI. Estimation

Here we describe estimators for TED and hence for MTTE. The estimators we provide here are not themselves new, being equivalent to estimators for treatment effects like those summarized in surveys such as Imbens and Wooldridge (2009) and Lee and Lemieux (2010). We therefore do not provide associated limiting distribution theory, since it corresponds to the standard theory of estimation of local polynomial estimators, albeit at boundary points. (See, e.g., Fan and Gijbels, 1996, and Porter, 2003.) What is new here is not the estimators themselves but their application to estimation of TED.[6]

First consider standard local linear estimation of sharp design RD models, having $T = T^* = I(X > c)$, with a uniform kernel. This is equivalent to selecting observations i such that $-\varepsilon \le X_i - c \le \varepsilon$ for a chosen bandwidth ε and then using just those observations to estimate the model,

$$Y_i = \alpha + (X_i - c)\,\beta + T_i^*\gamma_0 + (X_i - c)\,T_i^*\gamma_1 + e_i, \tag{13}$$

by ordinary least squares (OLS). Let $\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}_0$, and $\widehat{\gamma}_1$ be the OLS estimates of the coefficients α, β, γ_0, and γ_1. The line $(\alpha + \gamma_0) + (\beta + \gamma_1)(X - c)$ is a linear approximation to $E(Y \mid X, c \le X \le c + \varepsilon)$, so by the standard theory of local linear estimation, in the limit as $\varepsilon \to 0$ (at an appropriate rate), this approximation becomes equal to the tangent line to the function $g(x)$ as $x \longrightarrow c^+$, making $plim(\widehat{\alpha} + \widehat{\gamma}_0) = g_+(c)$ and $plim\left(\widehat{\beta} + \widehat{\gamma}_1\right) = g_+'(c)$. Similarly, the line $\alpha + \beta(X - c)$ is a linear approximation to $E(Y \mid X, c - \varepsilon \le X < c)$, making $plim(\widehat{\alpha}) = g_-(c)$ and $plim\left(\widehat{\beta}\right) = g_-'(c)$.

It follows immediately that

$$plim(\widehat{\gamma}_0) = g_+(c) - g_-(c) \quad \text{and}$$
$$plim(\widehat{\gamma}_1) = g_+'(c) - g_-'(c), \tag{14}$$

and therefore, by theorem 1, $\widehat{\tau}(c) = \widehat{\pi}(c) = \widehat{\gamma}_0$ is a consistent estimator of the sharp design treatment effect and $\widehat{\pi}'(c) = \widehat{\gamma}_1$ is a consistent estimator of the sharp design TED. Given local policy invariance, by corollary 1, $\widehat{\tau}'(c) = \widehat{\gamma}_1$ is a consistent estimator of the sharp design MTTE.

This local linear estimator of $\widehat{\tau}(c)$ is standard in the literature. What we are adding here is just the observation that these same commonly used local linear estimators also provide estimates of derivatives, which is all that we need to recover TED and hence MTTE.

Porter (2003) suggests using local polynomial regressions rather than ordinary kernel (locally constant) estimators because inclusion of the linear terms $(X_i - c)$ and $(X_i - c)\,T_i$ reduces small sample boundary bias in $\widehat{\alpha}$ and $\widehat{\gamma}_0$. For the same reason, to reduce bias in $\widehat{\beta}$ and $\widehat{\gamma}_1$, it could be advantageous to use local quadratic rather than local linear estimation, which with a uniform kernel corresponds to ordinary least squares estimation of the model:

$$Y_i = \alpha + (X_i - c)\,\beta + (X_i - c)^2\,\delta + T_i^*\gamma_0 + (X_i - c)\,T_i^*\gamma_1 + (X_i - c)^2\,T_i^*\gamma_2 + \widetilde{e}_i. \tag{15}$$

[6] Note that TED is a function of regression derivatives, which if estimated nonparametrically converge at slower rates than estimates of levels. So, for example, in equation (9) one can asymptotically ignore the estimation error in $\pi(c)$ when evaluating the distribution of $\tau(c_{new})$.

Equation (14) depends only on consistency of local (quadratic) polynomial estimation, and so will continue to hold if equation (15) is used instead of equation (13) for estimation.

For fuzzy designs, we can still use equation (13) or (15); in addition, we can use local linear or local quadratic estimation on the same observations i such that $-\varepsilon \leq X_i - c \leq \varepsilon$ to estimate

$$T_i = \alpha^T + (X_i - c)\,\beta^T + T_i^* \gamma_0^T + (X_i - c)\,T_i^* \gamma_1^T + e_i^T \tag{16}$$

or

$$T_i = \alpha^T + (X_i - c)\,\beta^T + (X_i - c)^2\,\delta^T + T_i^* \gamma_0^T + (X_i - c)\,T_i^* \gamma_1^T + (X_i - c)^2\,T_i^* \gamma_2^T + \tilde{e}_i^T \tag{17}$$

by ordinary least squares, yielding estimated coefficients including $\widehat{\gamma}_0^T$ and $\widehat{\gamma}_1^T$. By the exact same derivations as above, we have consistent estimators

$$plim\left(\widehat{\gamma}_0^T\right) = p\,(c) = f_+\,(c) - f_-\,(c) \quad \text{and}$$
$$plim\left(\widehat{\gamma}_1^T\right) = p'\,(c) = f_+'\,(c) - f_-'\,(c)\,. \tag{18}$$

Then by theorem 2, we have consistent estimators of the fuzzy design treatment effect and TED given by

$$\widehat{\pi}_f\,(c) = \widehat{\gamma}_0 / \widehat{\gamma}_0^T \quad \text{and} \quad \widehat{\pi}_f'\,(c) = \left(\widehat{\gamma}_1 - \widehat{\gamma}_1^T \widehat{\pi}_f\,(c)\right) / \widehat{\gamma}_0^T, \tag{19}$$

and, given policy invariance, the estimated fuzzy design MTTE is $\widehat{\tau}_f'\,(c) = \widehat{\pi}_f'\,(c)$.

In empirical practice, it is common to add other covariates Z_i as additional controls in these regressions. Adding or omitting these additional terms can be helpful for estimation precision but does not affect the consistency of the estimators described. All of our results extend immediately to the inclusion of covariates.[7]

VII. Empirical Application

We provide two empirical applications of our results. The first is a sharp design RD investigating the impact of the Adams Scholarship program on college choices, following Goodman (2008). The second is a fuzzy design RD assessing the impact of Superfund cleanups on housing prices, based on Greenstone and Gallagher (2008). Due to space limitations, we present this second empirical application in an online supplemental appendix.

The Adams Scholarship program is a U.S. merit-based scholarship program in Massachusetts. The program waives tuition at in-state public colleges if a student's score on a standardized test (the Massachusetts Comprehensive Assessment System, or MCAS) exceeds certain thresholds. The program is intended to attract talented students to the state's public colleges.

Applying a standard RD analysis, Goodman (2008) finds that qualifying for an Adams Scholarship induces 7.6% of recipients (at the threshold) to choose four-year public colleges instead of four-year private colleges. Goodman (2008) also uses a differences-in-differences (DID) analysis (comparing the cohorts immediately before and after the introduction of the scholarship program) and shows substantial treatment effect heterogeneity by academic skill levels. In particular, winners near the treatment threshold, who thus have relatively lower academic skills, respond much more strongly to the scholarship than more highly skilled winners. This is likely because highly skilled students can gain admission to private colleges of much higher quality than public colleges, and so the relatively small price reduction resulting from the Adams Scholarship is insufficient to compensate for the difference in school quality. In contrast, for the lowest-skilled winners (those with test scores right above the threshold), the quality difference is smaller or nonexistent, making the choice of a public college more worthwhile given its lower price.

Figure 2 shows the probability of choosing a four-year public college as a function of the number of grade points from the eligibility threshold. As is clear figure, the probability of choosing a four-year public college jumps substantially at the threshold but then declines quickly with test scores above the threshold. This dramatic downward slope change right above the threshold suggests that the Adams Scholarship mainly attracts relatively less skilled winners to public colleges, consistent with Goodman's DID analysis.

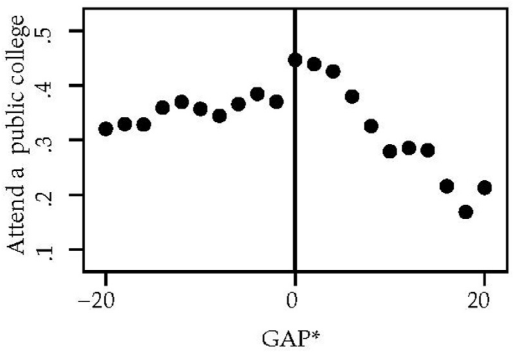
We formally investigate the hypothesis that the impact of an Adams Scholarship on the probability of choosing a public college declines as test scores rise by estimating TED. Further, we argue that local policy invariance plausibly holds in this application, making TED equal MTTE, which then allows us to estimate how the effects of the Adams Scholarship would change if the eligibility threshold were marginally changed. We show that this provides important information for public policy if the goal is to increase public college enrollment or college attendance in general.

The running variable X is the minimum distance to the relevant eligibility threshold, defined as the number of grade points by which a student succeeds or fails to win a scholarship.[8] In the data X, referred to as GAP^* in Goodman (2008), ranges from –132 to 20. The treatment indicator T is a dummy indicating whether a student is eligible for an Adams Scholarship. By construction, a student is eligible for the scholarship if and only if her or his test score is above the

[7] A conditional on covariates MTTE could be obtained by including covariates interacted with T^* in the estimating regressions, and an unconditional MTTE would then be obtained by averaging the conditional MTTE over covariate values.

[8] A student qualifies for Adams Scholarship if her or his total MCAS score (the sum of English and mathematics scores) falls in the top 25% of her or his school district grade distribution, so each school district has its own threshold. A student can also qualify if one of her or his scores reaches 260 and the other reaches 240. The running variable is therefore constructed as the minimum distance to the relevant threshold. Details about the construction of the running variable can be found in Good (2008).

Figure 2.—Probability of Attending a Four-Year College and Number of Grade Points from the Eligibility Threshold



threshold, so the RD model is a sharp design. We examine three outcomes: whether a student chooses a four-year public college, a four-year private college, or any college. Most of the students in the sample choose either a four-year public or a four-year private college (other choices include two-year colleges and trade schools).

We use the sample of the class of 2005 from Goodman (2008), the first cohort exposed to the scholarship program.[9] We estimate models like those in Goodman (2008). In particular, we estimate RD regression models with regressors X, the crossing threshold dummy $T^* = 1(X > 0)$, and an interaction term X times T^*, equivalent to local linear regressions above and below the threshold with a uniform kernel. In practice, different choices of kernel functions rarely make a difference. We consider two very different bandwidths to check the sensitivity of our results to greatly varying the weight put on observations close to the threshold. We start with the bandwidth used in Goodman (2008), which limits the sample to students with test scores within 20 points of the eligibility threshold. We then cut the bandwidth by half, using the sample of students with test scores within 10 points of the threshold.

Table 1 presents the estimation results. Estimates in the top panel are based on the bandwidth $|X| \leq 10$, and those in the bottom panel are based on $|X| \leq 20$. The corresponding sample sizes are 18,456 and 27,885, respectively. In each panel, the first row reports the estimated RD treatment effect at the current eligibility threshold, and the second row reports the estimated TED, which is also MTTE given local policy invariance. Based on MTTE, the last row in each panel reports what the estimated RD treatment effect would be if the threshold were marginally lowered by two grade points.[10] For each outcome, we report estimates both with and without controlling for covariates. The covariates are demographic variables consisting of indicators for female, black, Hispanic, poverty status, receiving special education or vocational education, limited English proficiency, speaking English as a second language, and coming from a medium-poverty district (if the poverty rate of the graduating class is between 20% and 40%) or a high-poverty district (if the poverty rate of the graduating class is above 40%). We report robust standard errors clustered at the school district level. Standard errors for the RD treatment effect at the new threshold are computed using the delta method.

As table 1 shows, the estimates are robust to different window widths used. Almost all of the estimates are significant at the 1% level. The point estimates barely change regardless of whether we control for covariates. The fact that the estimated TED (MTTE) is insensitive to including covariates provides strong evidence that the slope change at the cutoff is not driven by potentially omitted covariates, but rather indicates a direct relationship between students' response to the Adams Scholarship and test scores.

Consistent with Goodman's results, estimates in the top panel show that at the RD threshold, qualifying for an Adams Scholarship increases the probability of choosing a four-year public college by about 8% and decreases the probability of choosing a four-year private college by 7% to 8%. The impact on the overall college attendance is less than 2%. These results indicate that among marginal winners, the Adams Scholarship mainly encourages students to switch from four-year private colleges to four-year public colleges and has little impact on overall college attendance at the current eligibility threshold.

The estimated TED or MTTE for the four-year public college is –1.9%, which is significant at the 1% level, implying that the impact of Adams Scholarship negatively depends on the distance to the eligibility threshold, so students with higher test scores are less responsive, and if local policy invariance holds, then the Adams Scholarship program would encourage more students to choose four-year public colleges if the eligibility threshold were marginally lowered. The estimated TED for choosing a four-year private college is of a similar magnitude but has an opposite sign, implying that switching from private to public holds not only among the marginal winners at the current threshold but also among students with slightly lower or higher scores.

A sufficient (though stronger than necessary) condition for local policy invariance to hold in this application is if any given student's choice of college with an Adams Scholarship and her or his choice without one (one of these choices is a counterfactual) would not change if the eligibility threshold itself were marginally changed, ceteris paribus. Local policy invariance allows these choices to depend on the student's skill level and hence on her or his test score. Local policy invariance could be violated here if a marginal change in threshold changed people's perceptions of the value of attending a public college vs alternatives. So, for example, local policy invariance could be violated (through general equilibrium effects) if the marginally increased enrollment in public college due to a marginally increased number of offered scholarships caused a perceived decline in the value

[9] In his DID analysis, Goodman (2008) additionally uses data on the pretreatment cohort, the class of 2004.

[10] This is because the MCAS test score is in multiples of 2. Note also that since there are multiple thresholds, this threshold change means that all the relevant thresholds are lowered by two points.

TABLE 1.—RD ESTIMATES OF THE EFFECT OF ADAMS SCHOLARSHIP ON COLLEGE CHOICES

	Four-Year Public College		Four-Year Private College		Any College	
A. $\|X\| <= 10$						
Treatment effect	0.081	0.082	−0.080	−0.071	0.012	0.019
	(0.015)***	(0.015)***	(0.015)***	(0.015)***	(0.009)	(0.008)**
MTTE(TED)	−0.019	−0.019	0.018	0.017	−0.004	−0.004
	(0.003)***	(0.002)***	(0.002)***	(0.002)***	(0.001)***	(0.001)***
Treatment effect_new	0.120	0.120	−0.117	−0.105	0.019	0.027
	(0.015)***	(0.015)***	(0.015)***	(0.015)***	(0.009)**	(0.000)***
B. $\|X\| <= 20$						
Treatment effect	0.075	0.076	−0.061	−0.056	0.023	0.027
	(0.011)***	(0.011)***	(0.012)***	(0.011)***	(0.008)***	(0.006)***
MTTE(TED)	−0.017	−0.017	0.013	0.012	−0.003	−0.003
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Treatment effect_new	0.110	0.110	−0.086	−0.079	0.028	0.033
	(0.011)***	(0.011)***	(0.012)***	(0.011)***	(0.008)***	(0.006)***
Covariates	No	Yes	No	Yes	No	Yes

The sample size for the top panel A is 18,456, and for the bottom panel B is 27,885. Treatment effect - new refers to the RD treatment effect if the eligibility threshold were marginally lowered by two grade points. Robust standard errors are in parentheses. Significant at *10%, **5%, ***1%.

or prestige of the public college education. Or a violation might be possible through peer effects, for example, if seeing more of one's friends qualify for an Adams Scholarship changed one's own college choice. It seems unlikely that the magnitudes of these effects could be large enough to cause more than a very small difference between TED and MTTE. For example, the monetary value of the scholarship is relatively low,[11] making it unlikely to have a significant impact on the perceived value of the public schools.

Given local policy invariance, our estimates show that lowering the scholarship grade threshold by 1% would increase the probability of attending a four-year public college by 1.9% and would increase the probability of attending any college by 0.4%. Both estimates are significant at the 1% level. Intuitively, it seems plausible that Adams Scholarships may induce students with relatively low academic skills to go to college, since these students also tend to come from poor families (see more discussion on this in Goodman, 2008). Knowledge of these magnitudes should be useful for policymakers for assessing the likely impacts and costs of changing the scholarship eligibility requirements.

VIII. Conclusion

We have shown that in RD models, the treatment effect derivative (TED) is nonparametrically identified and easily estimated under smoothness assumptions that are already assumed in empirical applications of RD models.

This TED estimate can be useful in investigating treatment effect heterogeneity with respect to the running variable and in investigating external validity near the RD threshold. Having TED equal 0 is a necessary condition (which can be easily tested) for locally constant treatment effects. More generally, having TED be relatively large in magnitude may raise concern about the external validity of an RD LATE estimate, since a large TED indicates that a small change in the context, such as a small change in the running variable, is associated with large changes in average treatment effects.

We also show that if a local policy invariance assumption holds, then TED equals the marginal threshold treatment effect (MTTE), which can then be used to evaluate the likely changes in treatment effects that would result from a small change in the cutoff threshold. Local policy invariance is essentially a ceteris paribus assumption of the type often used in, for example, partial equilibrium analyses. Equivalently, local policy invariance is a type of external validity assumption for applying program evaluation results when the environment changes. In any particular application, one may assess from the institutional setting whether local policy invariance is likely to hold, at least approximately. To save space, some theoretical extensions of these results based on higher-order derivatives are provided in the online appendix to this paper.

We empirically apply our results first to evaluation of the Adams Scholarship program, where we find TED to be large and significant. We argue that this is a scenario where local policy invariance is at least a plausible approximation, so we can interpret TED as MTTE and hence provide counterfactual estimates regarding likely changes in the treatment effect that would result if the scholarship qualifying threshold were marginally raised or lowered.

We provide another empirical application in the online appendix. That application, a fuzzy design RD, considers estimation of the effects of Superfund cleanup eligibility on nearby housing prices. Greenstone and Gallagher (2008) found these effects to be quite small. In that application we find the estimated TED is also numerically small, implying that the impacts of eligibility on housing prices would remain small for sites with marginally higher or lower hazardous waste levels. Local policy invariance is also plausible in this application, as areas with hazardous waste sites are geographically separated and constitute a very small fraction of total available housing, so any potential general equilibrium effects should be negligible. Our MTTE then suggests that marginally raising or lowering the regulatory

[11] Adams Scholarships covered about 16% to 24% of the direct cost of attending public colleges in Massachusetts in 2005 according to Goodman (2008). In addition, the costs of these colleges were already relatively low.

threshold would not result in a significant change in the estimated effects of Superfund cleanups on nearby housing prices.

REFERENCES

Abbring, J. H., and J. J. Heckman, "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in J. J. Heckman & E. E. Leamer, eds., *Handbook of Econometrics* (New York: Elsevier, 2007).

Abdulkadiroglu, A., J. D. Angrist, and P. A. Pathak, "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica* 82 (2014), 137–196.

Angrist, J. D. and J.-S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton, NJ: Princeton University Press, 2008).

Angrist J., and M. Rokkanen, "Wanna Get Away? RD Identification Away from the Cutoff," NBER working paper 18662 (2012).

Card, D., D. S. Lee, Z. Pei, and A. Weber, "Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kink Design," NBER working paper 18564 (2012).

Carneiro, P., J. J. Heckman, and E. Vytlacil, "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin," *Econometrica* 78 (2010), 377–394.

Dinardo, J., and D. S. Lee, "Program Evaluation and Research Designs" (pp. 463–536), in *Handbook of Labor Economics* (New York: Elsevier, 2011).

Dong, Y., "Jump or Kink? Identification of Binary Treatment Regression Discontinuity Design without the Discontinuity," unpublished manuscript (2014a).

——— "An Alternative Assumption to Identify LATE in Regression Discontinuity Designs," unpublished manuscript (2014b).

Fan, J., and I. Gijbels, *Local Polynomial Modelling and Its Applications* (London: Chapman and Hall, 1996).

Goodman, J., "Who Merits Financial Aid? Massachusetts' Adams Scholarship," *Journal of Public Economics* 92 (2008), 2121–2131.

Greenstone, M., and J. Gallagher, "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program," *Quarterly Journal of Economics* 123 (2008), 951–1003.

Hahn, J., P. E. Todd, and W. Van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica* 69 (2001), 201–209.

Heckman, J. J., "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy," *Journal of Economic Literature* 48 (2010), 356–398.

Imbens, G. W., and T. Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics* 142 (2008), 615–635.

Imbens, G. W., and J. M. Wooldridge, "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature* 47 (2009), 5–86.

Jacob, B. A., and L. Lefgren, "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis," this REVIEW 86 (2004), 226–244.

Kirabo Jackson, C., "Do Students Benefit from Attending Better Schools? Evidence from Rule-Based Student Assignments in Trinidad and Tobago," *Economic Journal* 120 (2010), 1399–1429.

Lee, D. S., "Randomized Experiments from Non-Random Selection in U.S. House Elections," *Journal of Econometrics* 142 (2008), 675–697.

Lee, D. S., and T. Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature* 48 (2010), 281–355.

Maynard, R. A., K. A., Couch, C. Wing, and T. D. Cook, "Strengthening the Regression Discontinuity Design Using Additional Design Elements: A Within-Study Comparison," *Journal of Policy Analysis and Management* 32 (2013), 853–877.

Porter, J. R., "Estimation in the Regression Discontinuity Model," unpublished manuscript (2003).

Rubin, D. B., "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology* 66 (1974), 688–701.

Thistlethwaite, D., and D. Campbell, "Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment," *Journal of Educational Psychology* 51 (1960), 309–317.

APPENDIX

Proofs

Proof of theorem 1. Define $G_t(x) = E(Y(t) \mid X)$, so $\pi(c) = G_1(c) - G_0(c)$. Now

$$\begin{aligned} G_0(c) &= E(Y(0) \mid X = c) = \lim_{\varepsilon\to 0} E(Y(0) \mid X = c - \varepsilon) \\ &= \lim_{\varepsilon\to 0} E(Y(0) \mid X = c - \varepsilon, T = 0) \\ &= \lim_{\varepsilon\to 0} E(Y \mid X = c - \varepsilon) = g_-(c) \end{aligned}$$

and, similarly, $G_1(c) = g_+(c)$, which gives equation (1).

By assumption, $G_t(x)$ is continuously differentiable at $x = c$. This differentiability means that $G_0'(c)$ equals its own one-sided derivative $G_{0-}'(c)$. Similarly $G_0(c) = G_{0-}(c)$ and from above $G_0(c) = g_-(c)$. Using these equalities we have

$$\begin{aligned} G_0'(c) = G_{0-}'(c) &= \lim_{\varepsilon\to 0} \frac{G_0(c-\varepsilon) - G_0(c)}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{E(Y(0) \mid X = c-\varepsilon) - g(c)}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{E(Y(0) \mid X = c-\varepsilon, T = 0) - g(c)}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{E(Y \mid X = c-\varepsilon) - g(c)}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{g(c-\varepsilon) - g(c)}{\varepsilon} = g_-'(c). \end{aligned}$$

An analogous derivation using $x = c + \varepsilon$ gives $G_1'(c) = g_+'(c)$. Now $\pi(c) = G_1(c) - G_0(c)$, and taking the ordinary derivative of this expression and substituting in the above equalities gives

$$\pi'(c) = G_1'(c) - G_0'(c) = g_+'(c) - g_-'(c),$$

which proves the theorem.

Proof of theorem 2. By our definitions, $p(x) = \Pr(T(0) < T(1) \mid X = x)$, $g(x) = E(Y \mid X = x)$, $g_+(x) = \lim_{\varepsilon\to 0} g(x+\varepsilon)$ and $g_-(x) = \lim_{\varepsilon\to 0} g(x-\varepsilon)$. Note that in all these definitions and the derivation below, we hold the threshold fixed at $C = c$. For now, assume no defiers at x, a value near c, but since we will evaluate x actually at $x = c$, we only need no defiers holding at the threshold c.

$$\begin{aligned} & g_+(x) - g_-(x) \\ &= \lim_{\varepsilon\to 0} E(Y \mid X = x+\varepsilon) - \lim_{\varepsilon\to 0} E(Y \mid X = x-\varepsilon) \\ &= \lim_{\varepsilon\to 0} [E(Y(1) \mid X = x+\varepsilon, T(0) = T(1) = 1) \Pr(T(0) = T(1) \\ &\quad = 1 \mid X = x+\varepsilon)] \\ &\quad - \lim_{\varepsilon\to 0} [E(Y(1) \mid X = x-\varepsilon, T(0) = T(1) = 1) \Pr(T(0) = T(1) \\ &\quad = 1 \mid X = x-\varepsilon)] \\ &\quad + \lim_{\varepsilon\to 0} [E(Y(0) \mid X = x+\varepsilon, T(0) = T(1) = 0) \Pr(T(0) = T(1) \\ &\quad = 0 \mid X = x+\varepsilon)] \\ &\quad - \lim_{\varepsilon\to 0} [E(Y(0) \mid X = x-\varepsilon, T(0) = T(1) = 0) \Pr(T(0) = T(1) \\ &\quad = 0 \mid X = x-\varepsilon)] \\ &\quad + \lim_{\varepsilon\to 0} [E(Y(1) \mid X = x+\varepsilon, T(0) < T(1)) \Pr(T(0) \\ &\quad < T(1) \mid X = x+\varepsilon)] \\ &\quad - \lim_{\varepsilon\to 0} [E(Y(0) \mid X = x-\varepsilon, T(0) < T(1)) \Pr(T(0) \\ &\quad < T(1) \mid X = x-\varepsilon)] \\ &= [E(Y(1) \mid X = x, T(0) = T(1) = 1) \Pr(T(0) \\ &\quad = T(1) = 1 \mid X = x)] \\ &\quad - [E(Y(1) \mid X = x, T(0) = T(1) = 1) \Pr(T(0) \\ &\quad = T(1) = 1 \mid X = x)] \end{aligned}$$

$$
\begin{aligned}
&+ [E(Y(0) \mid X = x, T(0) = T(1) = 0) \Pr(T(0) \\
&\quad = T(1) = 0 \mid X = x)] \\
&- [E(Y(0) \mid X = x, T(0) = T(1) = 0) \Pr(T(0) \\
&\quad = T(1) = 0 \mid X = x)] \\
&+ [E(Y(1) \mid X = x, T(0) < T(1)) \Pr(T(0) < T(1) \mid X = x)] \\
&- [E(Y(0) \mid X = x, T(0) < T(1)) \Pr(T(0) < T(1) \mid X = x)] \\
&= E(Y(1) - Y(0) \mid X = x, T(0) < T(1)) \Pr(T(0) < T(1) \mid X = x) \\
&= \pi_f(x) p(x),
\end{aligned}
$$

where the second equality follows from monotonicity and hence no defiers, the third equality follows from smoothness of the probabilities of types and smoothness of conditional means of potential outcomes for each type of individual.

The above shows that $\pi_f(x) = \left[g_+(x) - g_-(x)\right] / p(x)$. With continuous differentiability, one can take the derivative of this expression with respect to x and evaluate the result at $x = c$ to get

$$\pi_f'(c) = \frac{g_+'(c) - g_-'(c)}{p(c)} - \pi_f(x) \frac{p'(c)}{p(c)}. \tag{A1}$$

Note that $p(x) = E(T(1) > T(0) \mid X = x) = E(T(1) - T(0) \mid X = x)$. Applying theorem 1, replacing Y with T, then gives $p(c) = f_+(c) - f_-(c)$ and $p'(c) = f_+'(c) - f_-'(c)$. Substituting these expressions into equation (A1) proves equation (6).