Extrapolating Treatment Effects in Multi-Cutoff Regression Discontinuity Designs
Supplemental Appendix

Matias D. Cattaneo[*]   Luke Keele[†]   Rocío Titiunik[‡]

Gonzalo Vazquez-Bare[§]

April 1, 2020

Abstract

This supplemental appendix includes a description of RD empirical papers with multiple cutoffs, further extensions and generalizations of our main methodological results, and omitted proofs and derivations. In addition, we outline an extension of our methods to the RD local randomization framework introduced by Cattaneo et al. (2015) and Cattaneo et al. (2017). Finally, R and Stata replication files are available at https://sites.google.com/site/rdpackages/replication/.

[*]Department of Operations Research and Financial Engineering, Princeton University.
[†]Department of Surgery and Biostatistics, University of Pennsylvania.
[‡]Department of Politics, Princeton University.
[§]Department of Economics, University of California at Santa Barbara.

Contents

SA-1 Empirical Examples of Multi-Cutoff RD Designs

Table SA-1 collects over 30 empirical papers in social, behavioral and biomedical sciences, where RD designs with multiple cutoffs are present. In most of these papers, the multi-cutoff feature of the RD designs was not exploited, but rather a pooling-and-normalizing approach was taken to conduct the empirical analysis (see Cattaneo et al., 2016, for methodological background).

Table SA-1: Empirical Papers Employing RD Designs with Multiple Cutoffs.

Citation	Place	Running Variable	Outcome Variable	Cutoffs
Abdulkadroglu et al. (2017)	US	Test scores	Test scores	Many
Angrist and Lavy (1999)	Israel	Cohort size	Test scores	2
Behaghel, Crépon and Sédillot (2008)	France	Age	Layoff rates	Many
Berk and de Leeuw (1999)	U.S.	Prison score	Re-conviction	4
Black, Galdo and Smith (2007)	U.S.	Training eligibility score	Job training and aid	Many
Brollo and Nannicini (2012)	Brazil	Population	Federal transfers	Many
Buddelmeyer and Skoufias (2004)	Mexico	Poverty score	Education attainment	Many
Canton and Blom (2004)	Mexico	Elgibility score	College outcomes	Many
Card and Shore-Sheppard (2004)	U.S.	Child age and income	Doctor visits	2
Card and Giuliano (2016)	US	Elgibility score	Test score	Many
Chay, McEwan and Urquiola (2005)	Chile	Elgibility score	School aid	13
Chen and Shapiro (2004)	U.S.	Prison score	Rearrest	5
Chen and Van der Klaauw (2008)	U.S.	Age	Disability awards	3
Clark (2009)	UK	Majority vote	Test scores	Many
Crost, Felter and Johnston (2014)	Phillipines	Poverty index	Conflict	22
Dell and Querubin (2017)	Vietnam	Geographic regions	Military strategy by location	Many
Dobkin and Ferreira (2010)	U.S.	Birthday	Education attainment	3
Edmonds (2004)	S. Africa	Age	Child outcomes	2 - 3
Garibaldi et al. (2012)	Italy	Income	Graduation	12
Goodman (2008)	U.S.	Test score	Scholarship	Many
Hjalmarsson (2009)	U.S.	Adjudication score	Re-conviction	2
Hoekstra (2009)	US	Test and grades	Earnings	Many
Hoxby (2000)	U.S.	Cohort size	Test scores	Many
Kane (2003)	U.S.	GPA	College attendance	Many
Klašnja and Titiunik (2017)	Brazil	Margin of victory	Incumbency advantage	Many
Kirkeboen, Leuven and Mogstad (2016)	Norway	Test scores	Earnings	Many
Litschig and Morrison (2013)	Brazil	Population	Poverty reduction	17
Spenkuch and Toniatti (2018)	US	County borders	Advertising and vote shares	Many
Snider and Williams (2015)	US	Distance from airports	Airfares	Many
Urquiola (2006)	Bolivia	Cohort size	Test scores	2
Urquiola and Verhoogen (2009)	Chile	Cohort size	Test scores	3
Van der Klaauw (2002)	U.S.	Aid score	Financial aid	Many
Van der Klaauw (2008)	U.S.	Poverty Score	School aid	Many

Note: "Many" refers to examples where either a large number of cutoff points are present or a continuum of cutoff points can be generated (e.g., the cutoff is a continuous random variable). This table excludes a large number of political science and related applications reported in Cattaneo et al. (2016, Supplemental Appenedix).

Table SA-2: Results for ACCES loan eligibility on Post-Education Enrollment in restricted ample

	Estimate	Bw	Eff. N	Robust BC Inference	
				p-value	95% CI
RD effects					
$C = -850$	0.061	85.0	85	0.282	[-0.052 , 0.179]
$C = -571$	0.169	136.3	133	0.103	[-0.039 , 0.428]
Weighted	0.121		218	0.056	[-0.003 , 0.274]
Pooled	0.073	161.5	307	0.254	[-0.046 , 0.173]
Naive difference					
$\mu_\ell(-650)$	0.728	239.4	440		
$\mu_h(-650)$	0.706	131.2	202		
Difference	0.021			0.634	[-0.050 , 0.081]
Bias					
$\mu_\ell(-850)$	0.560	58.0	57		
$\mu_h(-850)$	0.667	144.2	230		
Difference	-0.106			0.017	[-0.252 , -0.025]
Extrapolation					
$\tau_\ell(-650)$	0.128			0.022	[0.022 , 0.286]

Notes. Local polynomial regression estimation with MSE-optimal bandwidth selectors and robust bias corrected inference. See Calonico et al. (2014) and Calonico et al. (2018) for methodological details, and Calonico et al. (2017) and Cattaneo et al. (2020) for implementation. "Eff. N" indicates the effective sample size, that is, the sample size within the MSE-optimal bandwidth. "Bw" indicates the MSE-optimal bandwidth.

SA-2 Additional Empirical Results

In this section we explore the sensitivity of our empirical results when restricting the sample to the period 2007-2010. This check allows us to reduce the heterogeneity of the sample due to variation over time. The results are shown in Table SA-2. We find overall similar results, with a positive and significant extrapolated effect of 12.8 percentage points.

SA-3 Simulation Setup

We provide further details on the simulation setup used in the paper. Potential outcome regression functions are generated in the following way:

$$\mu_{0,h}(x) = \mathbf{r}_p(x)'\boldsymbol{\gamma}, \quad \mu_{0,\ell}(x) = \mu_{0,h}(x) + \Delta, \quad \mu_{1,c}(x) = \mu_{0,c}(x) + \tau$$

Table SA-3: Segments and Corresponding Parameters in Figure **??**

Segment	Parameter	Description
$\overline{ab}$	$\tau_\ell(\bar{x}) = \underbrace{\mu_{1,\ell}(\bar{x})}_{\text{Observable}} - \underbrace{\mu_{0,\ell}(\bar{x})}_{\text{Unobservable}}$	Extrapolation parameter of interest
$\overline{bc}$	$B(\bar{x}) = \underbrace{\mu_{0,\ell}(\bar{x})}_{\text{Unobservable}} - \underbrace{\mu_{0,\hbar}(\bar{x})}_{\text{Observable}}$	Control facing ℓ vs. control facing $\hbar$, at $X_i = \bar{x}$
$\overline{ac}$	$\tau_\ell(\bar{x}) + B(\bar{x}) = \underbrace{\mu_{1,\ell}(\bar{x})}_{\text{Observable}} - \underbrace{\mu_{0,\hbar}(\bar{x})}_{\text{Observable}}$	Treated facing ℓ vs. control facing $\hbar$, at $X_i = \bar{x}$
$\overline{ed}$	$B(\ell) = \underbrace{\mu_{0,\ell}(\ell)}_{\text{Observable}} - \underbrace{\mu_{0,\hbar}(\ell)}_{\text{Observable}}$	Control facing ℓ vs. control facing $\hbar$, at $X_i = \ell$

where $\mathbf{r}_p(x)$ is a p-th order polynomial basis for x. We set $p = 4$. The value of $\boldsymbol{\gamma}$, given below, is chosen by running a regression of the observed outcome for the high-cutoff on a fourth order polynomial using data from our empirical application. The observed variables are generated according to the following model:

$$X_i \sim \mathsf{Uniform}(-1000, -1)$$

$$\{C_i\}_{i=1}^N \sim \mathsf{FixedMargins}(N, N_\ell)$$

$$D_i = \mathbb{1}(X_i \geq C_i)$$

$$Y_i = \mu_{0,\hbar}(X_i) + \tau D_i + \Delta \mathbb{1}(C_i = \ell) + \mathsf{Normal}(0, \sigma^2)$$

where $\mathsf{FixedMargins}(N, N_\ell)$ denotes a fixed-margins distribution that assigns N_ℓ observations out of the total N sample to face cutoff ℓ and the remaining $N - N_\ell$ ones to face cutoff $\hbar$.

4

Proceeding as above, we set the parameters of the data generating process as follows:

$$\boldsymbol{\gamma} = (-14.089, -0.074, -1.372e(-4), -1.125e(-07), -3.444e(-11))'$$

$$\Delta = -0.14, \qquad \tau = 0.19, \qquad \sigma^2 = 0.3^2$$

$$N \in \{1000, 2000, 5000\}, \quad N_\ell = N/2$$

$$\ell = -850, \quad \hbar = -571, \quad \bar{x} = -650.$$

All these parameters were estimated using local polynomial and related methods.

SA-4 Extensions and Generalizations

We briefly discuss some extensions and generalizations of our main extrapolation results.

SA-4.1 Conditional-on-covariates Constant Bias

Following Abadie (2005), we can relax the constant bias assumption to hold conditionally on observable characteristics. Let the bias term conditional on a vector of observable covariates $Z_i = z \in \mathcal{Z}$ be denoted by $B(x,z) = \mathbb{E}[Y_i(0)|X_i = x, Z_i = z, C_i = \ell] - \mathbb{E}[Y_i(0)|X_i = x, Z_i = z, C_i = \hbar]$. Let $p(x,z) = \mathbb{P}(C_i = \ell|X_i = x, Z_i = z)$ denote the low-cutoff propensity score and $p(x) = \mathbb{P}(C_i = \ell|X_i = x)$.

We impose the following conditions:

Assumption SA-1 (Ignorable Constant Bias)

1. *Conditional bias: $B(\ell, z) = B(x, z)$ for all $x \in (\ell, \hbar)$ and for all $z \in \mathcal{Z}$.*

2. *Common support: $\delta < p(x, z) < 1 - \delta$ for some $\delta > 0$, $x \in \mathcal{X}$ and for all $z \in \mathcal{Z}$.*

3. *Continuity: $p(x, z)$ is continuous in x for all $z \in \mathcal{Z}$.*

Part 1 states that the selection bias is equal across cutoffs after conditioning on a covariates, part 2 is the usual assumption rulling out empty cells defined by the covariates Z_i, and part

3 assumes that the propensity score is continuous in the running variable. Then, letting $S_i = \mathbb{1}(C_i = \ell)$, we have the following result, which is proven in the supplemental appendix.

Theorem SA-1 (Covariate-Adjusted Extrapolation) *Under Assumption 1 and SA-1,*

$$
\begin{aligned}
\tau_\ell(\bar{x}) = {}& \mathbb{E}\left[\left.\frac{Y_i S_i}{p(\bar{x})}\right| X_i = \hbar\right] - \mathbb{E}\left[\left.\frac{Y_i(1-S_i)}{1-p(\bar{x}, Z_i)} \cdot \frac{p(\bar{x}, Z_i)}{p(\bar{x})}\right| X_i = x\right] \\
&+ \mathbb{E}\left[\left.\frac{Y_i(1-S_i)p(\bar{x}, Z_i)}{p(\bar{x})(1-p(\ell, Z_i))} \cdot \frac{f_{X|Z}(\hbar|Z_i)}{f_{X|Z}(\ell|Z_i)} \cdot \frac{f_X(\ell)}{f_X(\hbar)}\right| X_i = \ell\right] \\
&- \lim_{x \to \ell^-} \mathbb{E}\left[\left.\frac{Y_i S_i p(\bar{x}, Z_i)}{p(\bar{x})p(\ell, Z_i)} \cdot \frac{f_{X|Z}(\hbar|Z_i)}{f_{X|Z}(\ell|Z_i)} \cdot \frac{f_X(\ell)}{f_X(\hbar)}\right| X_i = x\right].
\end{aligned}
$$

This result is somewhat notationally convoluted, but it is straightforward to implement. It gives a precise formula for extrapolating RD treatment effects for values of the score above the cutoff ℓ, based on a conditional-on-Z_i constant bias assumption. All the unknown quantities in Theorem SA-1 can be replaced by consistent estimators thereof, under appropriate regularity conditions.

SA-4.2   Non-Constant Bias

Although the constant bias restriction (Assumption 2) is intuitive and allows for a helpful analogy with the difference-in-differences design, the RD setup leads for a natural extension via local polynomial extrapolation. The score is a continuous random variable and, under additional smoothness conditions of the bias function $B(x)$, we can replace the constant bias assumption with the assumption that $B(x)$ can be approximated by a polynomial expansion of $B(\ell)$ around $x \in (\ell, \hbar)$.

For example, using a polynomial of order one, we can approximate $B(x)$ at $\bar{x}$ as

$$
B(\bar{x}) \approx B(\ell) + \dot{B}(\ell) \cdot [\bar{x} - \ell] \tag{SA-1}
$$

where $\dot{B}(x) = \dot{\mu}_{0,\ell}(x) - \dot{\mu}_{0,\hbar}(x)$ and $\dot{\mu}_{d,c}(x) = \partial\mu_{d,c}(x)/\partial x$. This shows that the constant bias assumption $B(x) = B(\ell)$ can be seen as a special case of the above approximation, where

6

the first derivatives of $\mu_{0,\ell}(x)$ and $\mu_{0,\hbar}(x)$ are assumed equal to each other at $x = \bar{x}$. In contrast, the approximation in (SA-1) allows these derivatives to be different, and corrects the extrapolation at $\bar{x}$ using the difference between them at the point ℓ.

This idea allows, for instance, for different slopes of the control regression functions at $\bar{x}$, leading to $B(\bar{x}) \neq B(\ell)$. The linear adjustment in expression (SA-1) provides a way to correct the bias term to account for the difference in slopes at the low cutoff ℓ. This represents a generalization of the constant assumption, which allows the intercepts of $\mu_{0,\ell}(x)$ and $\mu_{0,\hbar}(x)$ to differ, but does not allow their difference to be a function of x. It is straightforward to extend this reasoning to employ higher order polynomials to approximate $B(\bar{x})$, at the cost of a stronger smoothness and extrapolation assumptions.

Assumption SA-2 (Polynomial-in-Score Bias)

1. *Smoothness:* $\mu_{d,c}(x) = \mathbb{E}[Y_i(d) \mid X_i = x, C_i = c]$ *are p-times continuously differentiable at $x = c$ for all $c \in \mathcal{C}$, $d = 0, 1$ and for some $p \in \{0, 1, 2, ...\}$.*

2. *Polynomial Approximation: there exists a $p \in \{0, 1, 2, ...\}$ such that, for $x \in (\ell, \hbar)$*

$$B(x) = \sum_{s=0}^{p} \frac{1}{s!} B^{(s)}(\ell) \cdot [x - \ell]^s$$

where $B^{(s)}(x) = \mu_{0,\ell}^{(s)}(x) - \mu_{0,\hbar}^{(s)}(x)$ and $\mu_{d,c}^{(s)}(x) = \partial^s \mu_{d,c}(x)/\partial x^s$.

The main extrapolation result in can be generalized as follows.

Theorem SA-2 *Under Assumption SA-2, for $\bar{x} \in (\ell, \hbar)$,*

$$\tau_\ell(\bar{x}) = \mu_\ell(\bar{x}) - \left[\mu_\hbar(\bar{x}) + \sum_{s=0}^{p} \frac{1}{s!} B^{(s)}(\ell) \cdot [x - \ell]^s \right].$$

This result establishes valid extrapolation of the RD treatment effect away from the low cutoff ℓ. This time the extrapolation is done via adjusting for not only the constant difference between the two control regression functions but also their higher-order derivatives.

7

Heuristically, this result justifies approximating the control regression functions by a higher-order polynomial, local to the cutoff, and then using the additional information about higher-derivatives to extrapolate the treatment effects.

SA-4.3 Many Cutoffs

All the identification results in the main paper hold for any number of cutoffs $J \geq 2$. The key issue to be assessed in this case is whether the constant bias assumption holds for all subpopulations. When this is the case, $\tau_c(x)$ is overidentified (for $x \geq c$), since there are many control groups that can be used to identify this parameter. In this setting, joint estimation can be performed using fixed effects models, as explained in the next section.

Furthermore, when more than two cutoffs are available, more control regression functions are therefore available for extrapolation. These could be combine to extrapolate or, alternatively, each of them could be used to extrapolate the RD treatment effect and only after these treatment effects could be combined. We relegate this interesting problem for future work.

SA-4.4 Multi-Scores and Geographic RD Designs

In many applications, RD designs include multiple scores (e.g., Papay et al., 2011; Reardon and Robinson, 2012; Keele and Titiunik, 2015). Examples include treatments assigned based on not exceeding a given threshold on two different scores (Figure SA-1(a)) or based on being one side of some generic boundary (Figure SA-1(b)). While these designs induce a continuum of cutoff points, it is usually better to analyze them using a finite number of cutoffs along the boundary determining treatment assignment.

For example, in Figure SA-1 we illustrate two settings with three chosen cutoff points (A, B, and C). In panel (a), the three cutoffs correspond to "extremes" over the boundary, while in panel (b) the cutoff points are chosen towards the "center". Once these cutoffs points are chosen, the analysis can proceed as discussed in the main paper. Usually, the

8

(a) Multi-Score RD Design

(b) Geographic RD Design
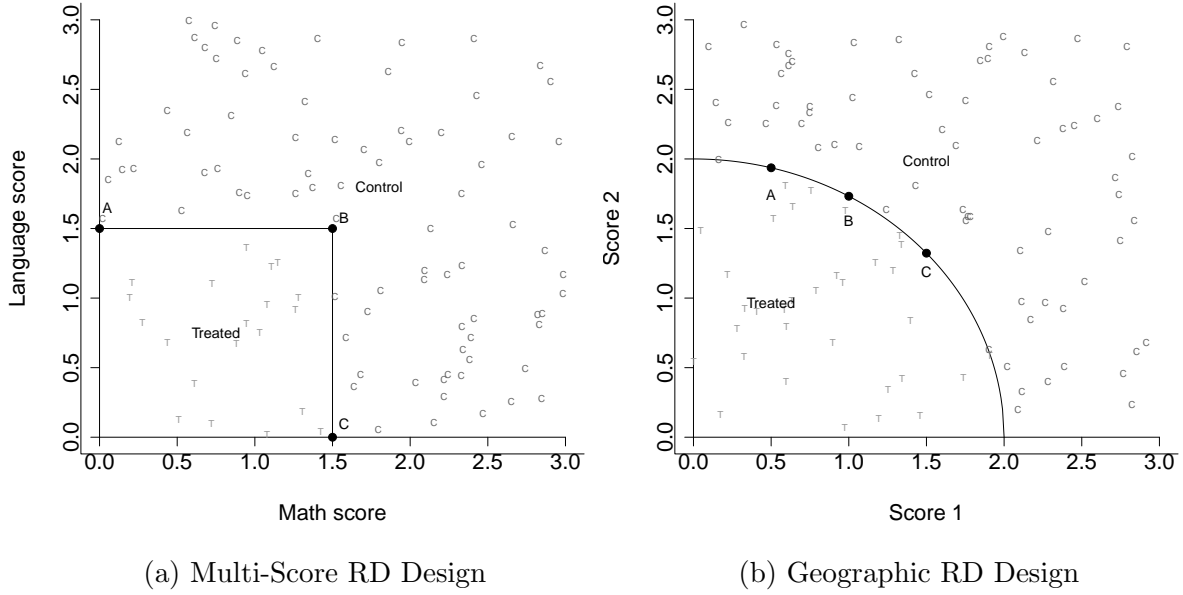
multidimensionality of the problem is reduced by relying on some metric that maps the multi-score feature of the design to a unidimensional problem. For instance, X_i is usually taken to be a measure of distance relative to the desired cutoff point. Once this mapping is constructed, all the ideas and methods discussed in the paper can be extended and applied to extrapolate RD treatment effects across cutoff points in multi-score RD designs.

SA-5 Relationship with Fixed Effects Models

Consider a separable model for the potential outcomes and only two cutoffs $\ell < \hbar$:

$$Y_i(0) = g_0(X_i) + \gamma_0 \mathbb{1}(C_i = \hbar) + \varepsilon_{0i}, \quad \mathbb{E}[\varepsilon_{0i} \mid X_i, C_i] = 0$$

$$Y_i(1) = g_1(X_i) + \gamma_1 \mathbb{1}(C_i = \hbar) + \varepsilon_{1i}, \quad \mathbb{E}[\varepsilon_{1i} \mid X_i, C_i] = 0$$

This model assumes that X_i, C_i and the error terms are separable. A key implication of separability is the absence of interaction between the cutoff and the score. In other words,

changing the cutoff only shifts the conditional expectation function without affecting the slope. The above model implies:

$$\tau_c(x) = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, C_i = c] = g_1(x) - g_0(x) + (\gamma_1 - \gamma_0)\mathbb{1}(c = \hbar)$$

Also, let $D_i = \mathbb{1}(X_i \geq C_i)$, $S_i = \mathbb{1}(C_i = \hbar)$, $\gamma \equiv \gamma_0$ and $\delta \equiv \gamma_1 - \gamma_0$, so that defining the observed outcome in the usual way, $Y_i = D_i Y_i(1) + (1 - D_i)Y_i(0)$, we get:

$$Y_i = g_0(X_i) + \gamma S_i + (g_1(X_i) - g_0(X_i))D_i + \delta D_i \times S_i + \varepsilon_i$$

where $\varepsilon_i = \varepsilon_{0i}D_i + \varepsilon_{1i}(1 - D_i)$ and $\mathbb{E}[\varepsilon_i \mid X_i, C_i] = 0$.

Although restrictive, this separable model nests several particular cases that are commonly used in RD estimation. For instance, if we assume:

$$g_0(x) = \alpha_0 + \beta_0 x, \quad g_1(x) = \alpha_1 + \beta_1 x,$$

the model reduces to:

$$Y_i = \alpha_0 + \beta_0 X_i + (\alpha_1 - \alpha_0)D_i + (\beta_1 - \beta_0)X_i \times D_i + \gamma S_i + \delta D_i \times S_i + \varepsilon_i$$

which is the usual linear model with an interaction between the score and the treatment, with two additional terms that account for the presence of two cutoffs.

The assumption of separability is sufficient for the selection bias to be constant across cutoffs. More precisely,

$$\mathbb{E}[Y_i(0) \mid X_i = x, C_i = \ell] = g_0(x)$$

$$\mathbb{E}[Y_i(0) \mid X_i = x, C_i = \hbar] = g_0(x) + \gamma_0$$

$$\Rightarrow B(x) = \gamma_0, \quad \forall x$$

In this setting, consider the case of many cutoffs is relatively straightforward. With $C_i \in \{c_0, c_1, ...c_J\} := \mathcal{C}$, the potential outcomes can be defined as:

$$Y_i(0) = g_0(X_i) + \sum_{j=1}^{J} \gamma_{0j} \mathbb{1}(C_i = c_j) + \varepsilon_i$$

$$Y_i(1) = g_{1j}(X_i) + \sum_{j=1}^{J} \gamma_{1j} \mathbb{1}(C_i = c_j) + \varepsilon_i$$

with observed outcome:

$$Y_i = g_0(X_i) + \sum_{j=1}^{J} \gamma_{0j} S_{ij} + (g_{1j}(X_i) - g_0(X_i))D_i + \sum_{j=1}^{J} \delta_j D_i \times S_{ij} + \varepsilon_i$$

where $S_{ij} = \mathbb{1}(C_i = c_j)$ and $\delta_j = \gamma_{1j} - \gamma_{0j}$. As before, the above model implies that the selection bias does not depend on the running variable:

$$B(x, c_j, c_k) = \gamma_{0j} - \gamma_{0k}$$

The equation for the observed outcome can be rewritten as:

$$Y_{ij} = \gamma_j + g(X_{ij}) + \theta_j D_{ij} + \tau_j(X_{ij})D_{ij} + \varepsilon_{ij}$$

where Y_{ij} is the outcome for unit i exposed to cutoff c_j and D_{ij} is equal to one if unit i exposed to cutoff c_j is treated, $D_{ij} = \mathbb{1}(X_i \geq c_j) \times \mathbb{1}(C_i = c_j)$. This model is similar to a one-way fixed effects model.

SA-5.1 Example: linear case

Using the above "fixed-effects notation", suppose g_0 and g_1 are linear:

$$Y_{ij}(0) = \gamma_{0j} + \beta_0 X_{ij} + \varepsilon_{ij}^0$$

$$Y_{ij}(1) = \gamma_{1j} + \beta_{1j} X_{ij} + \varepsilon_{ij}^1$$

where $Y_{ij}(d)$ is the potential outcome of unit i facing cutoff c_j under treatment d. As before, the fact that the slopes change with the treatment status but not with the cutoff is implied by separability between the score and the cutoff indicator. The observed outcome is:

$$Y_{ij} = \gamma_{0j} + \beta_0 X_{ij} + (\gamma_{1j} - \gamma_{0j})D_{ij} + (\beta_{1j} - \beta_0)X_{ij} \times D_{ij} + \varepsilon_{ij}$$

with $\varepsilon_{ij} = \varepsilon_{ij}^1 D_{ij} + \varepsilon_{ij}^0(1 - D_{ij})$.

Reparameterizing the model gives the estimating equation:

$$Y_{ij} = \gamma_j + \beta X_{ij} + \delta_j D_{ij} + \theta_j X_{ij} \times D_{ij} + \varepsilon_{ij}$$

which is a linear model including cutoff fixed effects, the running variable, the treatment variable with cutoff-varying coefficients and the interaction between the score and the treatment. In other words, this is the standard linear RD specification, but with different intercepts and slopes at each cutoff. Note that the coefficients for X_{ij} do not vary with j. This captures the restriction that the slopes of the conditional expectations under no treatment are the same across cutoffs and also leads to a straightforward specification test.

Under the linear specification, the treatment effect evaluated at c_k for units facing cutoff c_j with $c_k > c_j$ is given by:

$$\tau_{c_j}(\bar{x}) = \gamma_{1j} - \gamma_{0j} + (\beta_{1j} - \beta_0)\bar{x}$$

and can be estimated as:

$$\hat{\tau}_{c_j}(\bar{x}) = \hat{\delta}_j + \hat{\theta}_j \bar{x}$$

or simply as $\hat{\delta}_j$ if the running variable is appropriately re-centered.

Clearly, assuming a linear specification would in principle allow one to estimate the effects not only at the cutoffs but also at any other point in the range of the score. The treatment effects away from the cutoffs, however, are not identified nonparametrically and their identification relies purely on the functional form assumption, which can make them less credible.

The specification test mentioned above for the linear case consists on running the regression:

$$Y_{ij} = \gamma_j + \beta_j X_{ij} + \delta_j D_{ij} + \theta_j X_{ij} \times D_{ij} + \varepsilon_{ij}$$

and testing:

$$H_0: \quad \beta_1 = \beta_2 = ... = \beta_J.$$

SA-6 Local Randomization Methods

Our core extrapolation ideas can be adapted to the local randomization RD framework of Cattaneo et al. (2015) and Cattaneo et al. (2017). In this alternative framework, only units whose scores lay within a fixed (and small) neighborhood around the cutoff are considered, and their potential outcomes are regarded as fixed. The key source of randomness comes from the treatment assignment mechanism—the probability law placing units in control and treatment groups, and consequently the analysis proceeds as if the RD design was a randomized experiment within the neighborhood. See Rosenbaum (2010) and Imbens and Rubin (2015) for background on classical Fisher and Neyman approaches to the analysis of experiments.

As in the continuity-based approach, the Multi-Cutoff RD design can be analyzed using

local randomization methods by either normalizing-and-pooling all cutoffs or by studying each cutoff separately. However, to extrapolate away from an RD cutoff (i.e., outside the small neighborhood where local randomization is assumed to hold), further strong identifying assumptions are needed. To discuss these additional assumptions, we first formalize the local randomization (LR) framework for extrapolation in a Multi-Cutoff RD design.

Recall that $\mathcal{C} = \{\ell, \hbar\}$ for simplicity. Let $\mathcal{N}_x$ be a (non-empty) LR neighborhood around $x \in [\ell, \hbar]$, that is,

$$\mathcal{N}_x = [x - w, x + w], \qquad w > 0,$$

where the (half) window length may be different for each center point x. The other neighborhoods discussed below are defined analogously; for example, $\mathcal{N}_{\bar{x}} = [\bar{x} - w, \bar{x} + w]$ for some $w > 0$, possibly different across neighborhoods. Let also $y_{ic}(d, x)$ be a non-random potential outcome for unit i when facing cutoff c with treatment assignment d and running variable x. Consequently, in this Multi-Cutoff RD LR framework each unit has non-random potential outcomes $\{y_{i\ell}(0, x), y_{i\ell}(1, x), y_{i\hbar}(0, x), y_{i\hbar}(1, x)\}$, for each $x \in \mathcal{X}$. The observed outcome is $Y_i = y_{iC_i}(0, X_i)\mathbb{1}(X_i < C_i) + y_{iC_i}(1, X_i)\mathbb{1}(X_i \geq C_i)$ where $C_i \in \mathcal{C}$. Recall that our goal is to extrapolate the RD treatment effect to a point $\bar{x} \in (\ell, \hbar]$ on the support of the running variable.

Assumption SA-3 (LR Extrapolation)

(i) *For all i such that $X_i \in \mathcal{N}_\ell$,*

$$\{y_{i\ell}(0, x), y_{i\ell}(1, x), y_{i\hbar}(0, x)\} = \{y_{i\ell}(0, \ell), y_{i\ell}(1, \ell), y_{i\hbar}(0, \ell)\}$$

for all $x \in \mathcal{N}_\ell$, and are non-random. Furthermore, the treatment assignment mechanism is known.

(ii) *For $\bar{x} \in (\ell, \hbar]$ such that $\mathcal{N}_\ell \cap \mathcal{N}_{\bar{x}} = \emptyset$ and for all i such that $X_i \in \mathcal{N}_{\bar{x}}$,*

$$\{y_{i\ell}(0,x), y_{i\ell}(1,x), y_{i\hbar}(0,x)\} = \{y_{i\ell}(0,\bar{x}), y_{i\ell}(1,\bar{x}), y_{i\hbar}(0,\bar{x})\}$$

for all $x \in \mathcal{N}_{\bar{x}}$, and are non-random. Furthermore, the treatment assignment mechanism is known.

(iii) *There exists a constant $\Delta \in \mathbb{R}$ such that:*

$$y_{i\ell}(0,\ell) = y_{i\hbar}(0,\ell) + \Delta, \qquad \text{for all } i \text{ such that } X_i \in \mathcal{N}_\ell,$$

and

$$y_{i\ell}(0,\bar{x}) = y_{i\hbar}(0,\bar{x}) + \Delta, \qquad \text{for all } i \text{ such that } X_i \in \mathcal{N}_{\bar{x}}.$$

Assumption SA-3(i) is analogous to Assumption 1 in Cattaneo et al. (2015) applied to the RD cutoff $c = \ell$, except for the presence of one additional potential outcome, $y_{i\hbar}(0)$, which we will use for extrapolation in the Multi-Cutoff RD design. Likewise, Assumption SA-3(ii) postulates the existence of a LR neighborhood for the desired point of extrapolation $\bar{x}$. Finally, Assumption SA-3(iii) imposes a relationship between (the difference of) control potential outcomes in the LR neighborhood of $c = \ell$ and (the difference of) control potential outcomes in the LR neighborhood of $c = \bar{x}$, which we will use to impute the missing control potential outcomes for units exposed to the low cutoff $c = \ell$ but with scores within $\mathcal{N}_{\bar{x}}$. To conserve space and notation, we do not extend Assumption SA-3 to allow for regression adjustments within the LR neighborhoods as in Cattaneo et al. (2017), but we do include the corresponding results in our empirical application.

In the LR framework, extrapolation requires imputing both the assignment mechanism and the missing control potential outcomes $y_{i\ell}(0,\bar{x})$ within $\mathcal{N}_{\bar{x}}$. As a consequence, extrapolation beyond the standard LR neighborhood $\mathcal{N}_\ell$ requires very strong assumptions. Assumption SA-3 provides a set of conditions that lead to valid extrapolation. The parameter of interest

is the average effect of the treatment for units with $X_i \in \mathcal{n}_{\bar{x}}$:

$$\tau_{\mathsf{LR}} = \frac{1}{N_{\bar{x}}} \sum_{X_i \in \mathcal{n}_{\bar{x}}} (y_{i\ell}(1, \bar{x}) - y_{i\ell}(0, \bar{x}))$$

where $N_{\bar{x}}$ is the number of units inside the window $\mathcal{n}_{\bar{x}}$ around $\bar{x}$. Under Assumption SA-3, this parameter equals: $\frac{1}{N_{\bar{x}}} \sum_{X_i \in \mathcal{n}_{\bar{x}}} (y_{i\ell}(1, \bar{x}) - y_{i\hbar}(0, \bar{x})) - \Delta$, which is identifiable from the data. We implement this result as follows. First, we construct an estimate of Δ as the difference-in-means for control units facing cutoffs ℓ and $\hbar$ with $X_i \in \mathcal{n}_{\ell}$, which we denote by $\hat{\Delta}$:

$$\hat{\Delta} = \bar{Y}_{\ell}(0, \ell) - \bar{Y}_{\hbar}(0, \ell)$$

where

$$\bar{Y}_{\ell}(0, \ell) = \frac{1}{N_{\ell}^0(\ell)} \sum_{X_i \in \mathcal{n}_{\ell}} Y_i \mathbb{1}(C_i = \ell)(1 - D_i), \quad \bar{Y}_{\hbar}(0, \ell) = \frac{1}{N_{\hbar}(\ell)} \sum_{X_i \in \mathcal{n}_{\ell}} Y_i \mathbb{1}(C_i = \hbar)$$

and

$$N_{\ell}^0(\ell) = \sum_{X_i \in \mathcal{n}_{\ell}} \mathbb{1}(C_i = \ell)(1 - D_i), \quad N_{\hbar}(\ell) = \sum_{X_i \in \mathcal{n}_{\ell}} \mathbb{1}(C_i = \hbar).$$

Second, we estimate the treatment effects as:

$$\hat{\tau}_{\mathsf{LR}} = \bar{Y}_{\ell}(1, \bar{x}) - \bar{Y}_{\hbar}(0, \bar{x}) - \hat{\Delta}$$

where

$$\bar{Y}_{\ell}(1, \bar{x}) = \frac{1}{N_{\ell}(\bar{x})} \sum_{X_i \in \mathcal{n}_{\bar{x}}} Y_i \mathbb{1}(C_i = \ell), \quad \bar{Y}_{\hbar}(0, \bar{x}) = \frac{1}{N_{\hbar}(\bar{x})} \sum_{X_i \in \mathcal{n}_{\bar{x}}} Y_i \mathbb{1}(C_i = \hbar)$$

and

$$N_{\ell}(\bar{x}) = \sum_{X_i \in \mathcal{n}_{\bar{x}}} \mathbb{1}(C_i = \ell), \quad N_{\hbar}(\bar{x}) = \sum_{X_i \in \mathcal{n}_{\bar{x}}} \mathbb{1}(C_i = \hbar).$$

Finally, for the assignment mechanism, we assume:

$$\mathbb{P}[C_i = \hbar | X_i \in \mathcal{N}_c] = \frac{N_\hbar(c)}{N_\ell(c) + N_\hbar(c)}, \quad \forall i : X_i \in \mathcal{N}_c, c \in \{\ell, \bar{x}\}$$

and

$$\mathbb{P}[D_i = 0 | C_i = \ell, X_i \in \mathcal{N}_\ell] = \frac{N_\ell^0(\ell)}{N_\ell(\ell)}, \quad N_\ell(\ell) = \sum_{X_i \in \mathcal{N}_\ell} \mathbb{1}(C_i = \ell)$$

It is straightforward to show that under this assignment mechanism and Assumption SA-3, $\hat{\Delta}$ and $\hat{\tau}$ are unbiased for their corresponding parameters. Our approach is not the only way to develop LR methods for extrapolation, but for simplicity we focus on the above construction which mimics closely the continuity-based proposed in the previous sections.

In this setting, inference can be conducted using Fisherian randomization inference by permuting the cutoff indicator $\mathbb{1}(C_i = \ell)$ on the adjusted outcomes $Y_i^A = Y_i + \Delta \mathbb{1}(C_i = \hbar)$ among units in $\mathcal{N}_{\bar{x}}$. However, the inference procedure needs to account for the fact that Δ is unknown and needs to be estimated. We propose two alternatives to deal with this issue. The first one, suggested by Berger and Boos (1994), consists on constructing a $(1 - \eta)$-level confidence interval for Δ, S_η, and defining the p-value:

$$p^*(\eta) = \sup_{\Delta \in S_\eta} p(\Delta) + \eta$$

which can be shown to be valid in the sense that $\mathbb{P}[p^*(\eta) \leq \alpha] \leq \alpha$.

Our second inference procedure, based on Neyman's sampling approach, consists on using the standard normal distribution to approximate the distribution of the studentized statistic:

$$T = \frac{\hat{\tau}}{\sqrt{V_1 + V_{\hat{\Delta}}}}$$

where V_1 is the estimated variance of the difference in means $\bar{Y}_\ell(1, \bar{x}) - \bar{Y}_\hbar(0, \bar{x})$ and $V_{\hat{\Delta}}$ is the estimated variance of $\hat{\Delta}$.

SA-6.1 Empirical Application

We use our proposed LR methods to investigate the external validity of the ACCES program RD effects. As mentioned in the paper, our sample has observations exposed to two cutoffs, $\ell = -850$ and $\hbar = -571$. We begin by extrapolating the effect to the point $\bar{x} = -650$; our focus is thus the effect of eligibility for ACCES on whether the student enrolls in a higher education program for the subpopulation exposed to cutoff 850 when their SABER 11 score is 650.

Table SA-4 presents empirical results using the local randomization framework. We construct the neighborhoods $\mathcal{N}_\ell$ and $\mathcal{N}_{\bar{x}}$ using the 50 closest observations to the evaluation point of interest. To calculate $p^*(\eta)$, we construct a 99 percent confidence interval for Δ based on the normal approximation, which can be justified using large sample approximations in either a fixed potential outcomes model (Neyman) or a standard repeated sampling model (superpopulation). We estimate $\hat{\tau}_{\mathsf{LR}}$ using a constant model and using a linear adjustment (see Cattaneo et al., 2017, for details). Overall, the results are very similar to the ones obtained using the continuity-based approach. We find positive effects of around 20 percentage points that are significant at the 5 percent level using either Fisherian-based or Neyman-based inference.

To assess robustness of the LR methods, Figure SA-2 shows how the estimated effect and its corresponding randomization inference p-value change when varying the number of nearest neighbors used to construct $\mathcal{N}_\ell$ and $\mathcal{N}_{\bar{x}}$. The magnitude of the estimated effect remains stable when increasing the length of the window, particularly for the linear adjustment case which can help to reduce bias when the corresponding regression functions are not constant. In terms of inference, while the p-values we construct can be very conservative, we find significant effects at the 5 percent level when using around 45 observations in each neighborhood.

Table SA-4: Empirical Results under Local Randomization

	Window	Eff. N	Constant			Linear		
			Estimate	Fisher p	Neyman p	Estimate	Fisher p	Neyman p
$X_i \in n_\ell$								
$\bar{Y}_\ell(0,\ell)$	[-900 , -850)	50	0.502			0.527		
$\bar{Y}_h(0,\ell)$	[-881 , -817]	50	0.706			0.707		
Δ		100	-0.204	0.000	0.000	-0.180	0.000	0.021
$X_i \in n_{\bar{x}}$								
$\bar{Y}_\ell(1,\bar{x})$	[-675 , -626]	50	0.760			0.759		
$\bar{Y}_h(0,\bar{x})$	[-675 , -625]	50	0.743			0.743		
Diff		100	0.017	0.702	0.698	0.016	0.718	0.716
$\hat{\tau}_{\mathsf{LR}}$			0.220	0.042	0.001	0.196	0.037	0.029

Notes. Estimated effect under the local randomization framework. Randomization inference p-values for $\hat{\tau}_{\mathsf{LR}}$ constructed using the Berger and Boos (1994) method. Neyman-based p-values constructed using a large-sample normal approximation. Estimates calculated using a constant model based on difference in meansand a linear regression adjustment (see Cattaneo et al., 2017, for details). "Eff. N" indicates the effective sample size, that is, the sample size within the local randomization neighborhood.

Figure SA-2: Sensitivity Analysis for Local Randomization

(a) Constant Model   (b) Linear Adjustment
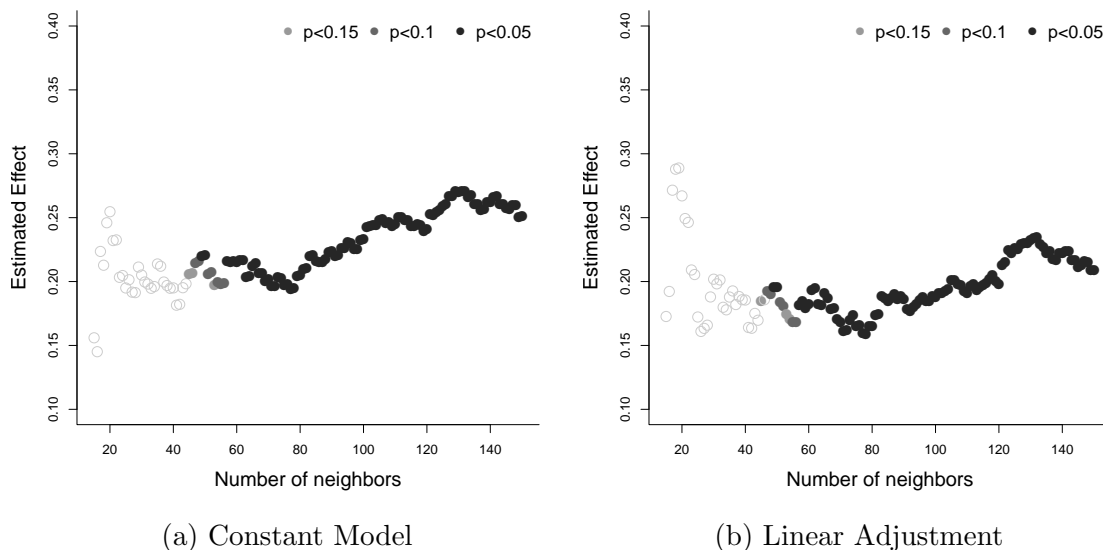
Notes. The figure plots the estimated effect as a function of the number of nearest neighbors used around the cutoff for estimation. The left panel plots the estimates under a constant model. The right panel plots the estimates using a linear adjustment model. Hollow markers indicate p-value ≥ 0.15. Light gray markers indicate p-value < 0.15. Dark gray markers indicate p-value < 0.1. Black markers indicate p-value < 0.05. Randomization inference p-values constructed using the Berger and Boos (1994) method.

SA-7 Proofs and Derivations

SA-7.1 Proof of Theorem 1

Follows immediately under the stated assumptions by the derivation provided in the paper.

SA-7.2 Proof of Theorem 2

We omit the i subscript to simplify notation. The average observed outcomes are:

$$\mathbb{E}[Y|X = x, C = c] = \mathbb{E}[Y(1)D(x,c) + Y(0)(1 - D(x,c))|X = x, C = c]$$
$$= \mathbb{E}[(Y(1) - Y(0))D(x,c)|X = x, C = c] + \mathbb{E}[Y(0)|X = x, C = c]$$

The difference in outcomes for units facing cutoffs $l < h$ at $X = x \in (l, h)$ is:

$$\Delta(x) = \mathbb{E}[Y|X = x, C = l] - \mathbb{E}[Y|X = x, C = h]$$
$$= \mathbb{E}[(Y(1) - Y(0))D(x,l)|X = x, C = l] - \mathbb{E}[(Y(1) - Y(0))D(x,h)|X = x, C = h]$$
$$+ \mathbb{E}[Y(0)|X = x, C = l] - \mathbb{E}[Y(0)|X = x, C = h]$$

Assuming parallel regression functions under no treatment, the double difference recovers:

$$\Delta(x) - \Delta(l)$$
$$= \mathbb{E}[Y|X = x, C = l] - \mathbb{E}[Y|X = x, C = h] - \lim_{x \uparrow l} \mathbb{E}[Y|X = x, C = l] + \mathbb{E}[Y|X = l, C = h]$$
$$= \mathbb{E}[(Y(1) - Y(0))D(x,l)|X = x, C = l] - \mathbb{E}[(Y(1) - Y(0))D^-(l,l)|X = l, C = l]$$
$$- \{\mathbb{E}[(Y(1) - Y(0))D(x,h)|X = x, C = h] - \mathbb{E}[(Y(1) - Y(0))D(l,h)|X = l, C = h]\}$$

where $D^-(l,l) = \lim_{x \uparrow l} D(x,l)$. Note that in a sharp design, $D(x,l) = 1$ and $D(x,h) = D(l,h) = D^-(l,l) = 0$ giving our previous result. After adding and subtracting,

$$\Delta(x) - \Delta(l)$$

$$= \mathbb{E}[Y|X = x, C = l] - \mathbb{E}[Y|X = x, C = h] - \lim_{x \uparrow l} \mathbb{E}[Y|X = x, C = l] + \mathbb{E}[Y|X = l, C = h]$$

$$= \mathbb{E}[(Y(1) - Y(0))(D(x,l) - D^-(l,l))|X = x, C = l]$$

$$+ \mathbb{E}[(Y(1) - Y(0))D^-(l,l)|X = x, C = l] - \mathbb{E}[(Y(1) - Y(0))D^-(l,l)|X = l, C = l]$$

$$- \{\mathbb{E}[(Y(1) - Y(0))D(x,h)|X = x, C = h] - \mathbb{E}[(Y(1) - Y(0))D(l,h)|X = l, C = h]\}$$

Under one-sided non-compliance, units below the cutoff never get the treatment, so $D(x,h) = D(l,h) = D^-(l,l) = 0$. In this case,

$$\Delta(x) - \Delta(l) = \mathbb{E}[(Y(1) - Y(0))D(x,l)|X = x, C = l]$$

$$= \mathbb{E}[Y(1) - Y(0)|X = x, C = l, D(x,l) = 1]\mathbb{P}[D(x,l) = 1|X = x, C = l]$$

It follows that

$$\frac{\Delta(x) - \Delta(l)}{\mathbb{E}[D|X = x, C = l]} = \mathbb{E}[Y(1) - Y(0)|X = x, C = l, D(x,l) = 1]$$

which in this case equals the (local) average effect on the compliers.

SA-7.3 Proof of Theorem SA-1

Following analogous steps to the derivation for the unconditional case, we obtain that:

$$\mathbb{E}[\tau_i|X_i = \bar{x}, Z_i = z, C_i = \ell] = \mathbb{E}[Y_i|X_i = \bar{x}, Z_i = z, C_i = \ell]$$

$$- \mathbb{E}[Y_i|X_i = \bar{x}, Z_i = z, C_i = \hbar] - B(\ell, z)$$

and

$$\tau_\ell(\bar{x}) = \mathbb{E}\left\{\mathbb{E}[\tau_i|X_i = \bar{x}, Z_i, C_i = \ell]|X_i = \bar{x}, C_i = \ell\right\}$$

Define $p(x, z) = \mathbb{P}(C_i = \ell|X_i = x, Z_i = z)$ and $S_i = \mathbb{1}(C_i = \ell)$. We have that:

$$\mathbb{E}[Y_i|X_i = \bar{x}, C_i = \ell, Z_i] = \mathbb{E}\left[\left.\frac{Y_i S_i}{p(\bar{x}, Z_i)}\right| X_i = \bar{x}, Z_i\right]$$

$$\mathbb{E}[Y_i|X_i = \ell, C_i = \hbar, Z_i] = \mathbb{E}\left[\left.\frac{Y_i(1 - S_i)}{1 - p(\ell, Z_i)}\right| X_i = \ell, Z_i\right]$$

and similarly for the remaining two terms, we obtain, assuming that $p(x, z)$ is continuous in x (and limits can be interchanged),

$$\mathbb{E}[\tau_i|X_i = \bar{x}, Z_i = z, C_i = \ell] = \mathbb{E}\left[\left.\frac{Y_i S_i}{p(\bar{x}, Z_i)}\right| X_i = \bar{x}, Z_i\right] - \mathbb{E}\left[\left.\frac{Y_i(1 - S_i)}{1 - p(\bar{x}, Z_i)}\right| X_i = \bar{x}, Z_i\right]$$

$$+ \mathbb{E}\left[\left.\frac{Y_i(1 - S_i)}{1 - p(\ell, Z_i)}\right| X_i = \ell, Z_i\right] - \lim_{x \to \ell^-} \mathbb{E}\left[\left.\frac{Y_i S_i}{p(\ell, Z_i)}\right| X_i = x, Z_i\right]$$

To simplify the notation, let $\tau_c(x, z) = \mathbb{E}[\tau_i|X_i = x, C_i = c, Z_i = z]$, $\tau_c(x) = \mathbb{E}[\tau_i|X_i = x, C_i = c]$ and $p(x) = \mathbb{P}(C_i = \ell|X_i = x)$. By Bayes' rule:

$$\tau_\ell(\bar{x}) = \int \tau_\ell(\bar{x}, z) dF_{Z|X,C}(z|\bar{x}, \ell)$$

$$= \int \tau_\ell(\bar{x}, z) \frac{p(\bar{x}, z)}{p(\bar{x})} dF_{Z|X}(z|\bar{x})$$

$$= \mathbb{E}\left[\left.\tau_\ell(\bar{x}, Z_i) \frac{p(\bar{x}, Z_i)}{p(\bar{x})}\right| X_i = \bar{x}\right]$$

Now replace $\tau_\ell(\bar{x}, Z_i)$ with its observed counterpart and split the outer expectation into the four summands. We have that:

$$\mathbb{E}\left[\left.\mathbb{E}\left[\left.\frac{Y_i S_i}{p(\bar{x}, Z_i)}\right| X_i = \bar{x}, Z_i\right] \frac{p(\bar{x}, Z_i)}{p(\bar{x})}\right| X_i = \bar{x}\right] = \mathbb{E}\left[\left.\mathbb{E}\left[\left.\frac{Y_i S_i}{p(\bar{x})}\right| X_i = \bar{x}, Z_i\right]\right| X_i = \bar{x}\right]$$

$$= \mathbb{E}\left[\left.\frac{Y_i S_i}{p(\bar{x})}\right| X_i = \bar{x}\right]$$

and

$$\mathbb{E}\left[\mathbb{E}\left[\left.\frac{Y_i(1-S_i)}{1-p(\ell,Z_i)}\right|X_i=\ell,Z_i\right]\frac{p(\bar{x},Z_i)}{p(\bar{x})}\right|X_i=\bar{x}\right]$$

$$=\mathbb{E}\left[\mathbb{E}\left[\left.\frac{Y_i(1-S_i)}{1-p(\ell,Z_i)}\cdot\frac{p(\bar{x},Z_i)}{p(\bar{x})}\right|X_i=\ell,Z_i\right]\right|X_i=\bar{x}\right]$$

$$=\mathbb{E}\left[\left.\frac{Y_i(1-S_i)}{p(\bar{x})}\cdot\frac{p(\bar{x},Z_i)}{1-p(\ell,Z_i)}\cdot\frac{f_{Z|X}(Z_i|\bar{x})}{f_{Z|X}(Z_i|\ell)}\right|X_i=\ell\right]$$

$$=\mathbb{E}\left[\left.\frac{Y_i(1-S_i)p(\bar{x},Z_i)}{p(\bar{x})(1-p(\ell,Z_i))}\cdot\frac{f_{X|Z}(\bar{x}|Z_i)}{f_{X|Z}(\ell|Z_i)}\cdot\frac{f_X(\ell)}{f_X(\bar{x})}\right|X_i=\ell\right]$$

Similarly,

$$\mathbb{E}\left[\mathbb{E}\left[\left.\frac{Y_i(1-S_i)}{1-p(\bar{x},Z_i)}\right|X_i=\bar{x},Z_i\right]\frac{p(\bar{x},Z_i)}{p(\bar{x})}\right|X_i=\bar{x}\right]=\mathbb{E}\left[\left.\frac{Y_i(1-S_i)}{1-p(\bar{x},Z_i)}\cdot\frac{p(\bar{x},Z_i)}{p(\bar{x})}\right|X_i=x\right]$$

and finally, under regularity conditions to interchange limits and expectations,

$$\mathbb{E}\left[\lim_{x\to\ell^-}\mathbb{E}\left[\left.\frac{Y_iS_i}{p(\ell,Z_i)}\right|X_i=x,Z_i\right]\frac{p(\bar{x},Z_i)}{p(\bar{x})}\right|X_i=\bar{x}\right]$$

$$=\lim_{x\to\ell^-}\mathbb{E}\left[\mathbb{E}\left[\left.\frac{Y_iS_i\cdot p(\bar{x},Z_i)}{p(\bar{x})p(\ell,Z_i)}\right|X_i=x,Z_i\right]\right|X_i=\bar{x}\right]$$

$$=\lim_{x\to\ell^-}\mathbb{E}\left[\left.\frac{Y_iS_ip(\bar{x},Z_i)}{p(\bar{x})p(\ell,Z_i)}\cdot\frac{f_{Z|X}(Z_i|\hbar)}{f_{Z|X}(Z_i|\ell)}\right|X_i=x\right]$$

$$=\lim_{x\to\ell^-}\mathbb{E}\left[\left.\frac{Y_iS_ip(\bar{x},Z_i)}{p(\bar{x})p(\ell,Z_i)}\cdot\frac{f_{X|Z}(\hbar|Z_i)}{f_{X|Z}(\ell|Z_i)}\cdot\frac{f_X(\ell)}{f_X(\hbar)}\right|X_i=x\right]$$

Putting all the results together,

$$\tau_\ell(\bar{x})=\mathbb{E}\left[\left.\frac{Y_iS_i}{p(\bar{x})}\right|X_i=\hbar\right]-\mathbb{E}\left[\left.\frac{Y_i(1-S_i)}{1-p(\bar{x},Z_i)}\cdot\frac{p(\bar{x},Z_i)}{p(\bar{x})}\right|X_i=x\right]$$

$$+\mathbb{E}\left[\left.\frac{Y_i(1-S_i)p(\bar{x},Z_i)}{p(\bar{x})(1-p(\ell,Z_i))}\cdot\frac{f_{X|Z}(\hbar|Z_i)}{f_{X|Z}(\ell|Z_i)}\cdot\frac{f_X(\ell)}{f_X(\hbar)}\right|X_i=\ell\right]$$

$$-\lim_{x\to\ell^-}\mathbb{E}\left[\left.\frac{Y_iS_ip(\bar{x},Z_i)}{p(\bar{x})p(\ell,Z_i)}\cdot\frac{f_{X|Z}(\hbar|Z_i)}{f_{X|Z}(\ell|Z_i)}\cdot\frac{f_X(\ell)}{f_X(\hbar)}\right|X_i=x\right].$$

SA-7.4   Proof of Theorem SA-2

Follows immediately under the stated assumptions by the derivation provided in the paper.

References

Abadie, A. (2005), "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, 72, 1–19.

Abdulkadroglu, A., Angrist, J. D., Narita, Y., Pathak, P. A., and Zarate, R. A. (2017), "Regression Discontinuity in Serial Dictatorship: Achievement Effects at Chicago's Exam Schools," *American Economic Review: Papers & Proceedings*, 107, 240–245.

Angrist, J. D., and Lavy, V. (1999), "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement*," *The Quarterly journal of economics*, 114, 533–575.

Behaghel, L., Crépon, B., and Sédillot, B. (2008), "The perverse effects of partial employment protection reform: The case of French older workers," *Journal of Public Economics*, 92, 696–721.

Berger, R. L., and Boos, D. D. (1994), "P Values Maximized Over a Confidence Set for the Nuisance Parameter," *Journal of the American Statistical Association*, 89, 1012–1016.

Berk, R. A., and 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, 1045–1052.

Black, D. A., Galdo, J., and Smith, J. A. (2007), "Evaluating the worker profiling and reemployment services system using a regression discontinuity approach," *The American Economic Review*, 97, 104–107.

Brollo, F., and Nannicini, T. (2012), "Tying Your Enemy's Hands in Close Races: The Politics of Federal Transfers in Brazil," *American Political Science Review*, 106, 742–761.

Buddelmeyer, H., and Skoufias, E. (2004), *An evaluation of the performance of regression discontinuity design on PROGRESA*, Vol. 827, World Bank Publications.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018), "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," *Journal of the American Statistical Association*, 113, 767–779.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017), "`rdrobust`: Software for Regression Discontinuity Designs," *Stata Journal*, 17, 372–404.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014), "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326.

Canton, E., and Blom, A. (2004), "Can Student Loans Improve Accessibility to Higher Education and Student Performance? An Impact Study of the Case of SOFES, Mexico," World Bank Working Paper 3425.

Card, D., and Giuliano, L. (2016), "Can tracking raise the test scores of high-ability minority students?" *American Economic Review*, 106, 2783–2816.

Card, D., and Shore-Sheppard, L. D. (2004), "Using discontinuous eligibility rules to identify the effects of the federal medicaid expansions on low-income children," *Review of Economics and Statistics*, 86, 752–766.

Cattaneo, M. D., Frandsen, B., and Titiunik, R. (2015), "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate," *Journal of Causal Inference*, 3, 1–24.

Cattaneo, M. D., Keele, L., Titiunik, R., and Vazquez-Bare, G. (2016), "Interpreting Regression Discontinuity Designs with Multiple Cutoffs," *Journal of Politics*, 78, 1229–1248.

Cattaneo, M. D., Titiunik, R., and Vazquez-Bare, G. (2017), "Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality," *Journal of Policy Analysis and Management*, 36, 643–681.

——— (2020), "Analysis of Regression Discontinuity Designs with Multiple Cutoffs or Multiple Scores," working paper.

Chay, K. Y., McEwan, P. J., and Urquiola, M. (2005), "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools," *The American Economic Review*, 95, 1237–1258.

Chen, M. K., and Shapiro, J. M. (2004), *Does Prison Harden Inmates?: A Discontinuity-based Approach*, Cowles Foundation for Research in Economics.

Chen, S., and Van der Klaauw, W. (2008), "The work disincentive effects of the disability insurance program in the 1990s," *Journal of Econometrics*, 142, 757–784.

Clark, D. (2009), "The performance and competitive effects of school autonomy," *Journal of Political Economy*, 117, 745–783.

Crost, B., Felter, J., and Johnston, P. (2014), "Aid under fire: Development projects and civil conflict," *American Economic Review*, 104, 1833–56.

Dell, M., and Querubin, P. (2017), "Nation building through foreign intervention: Evidence from discontinuities in military strategies," *The Quarterly Journal of Economics*, 1, 64.

Dobkin, C., and Ferreira, F. (2010), "Do school entry laws affect educational attainment and labor market outcomes?" *Economics of Education Review*, 29, 40–54.

Edmonds, E. V. (2004), "Does illiquidity alter child labor and schooling decisions? evidence from household responses to anticipated cash transfers in South Africa," National Bureau of Economic Research Working Paper No 10265.

Garibaldi, P., Giavazzi, F., Ichino, A., and Rettore, E. (2012), "College cost and time to complete a degree: Evidence from tuition discontinuities," *Review of Economics and Statistics*, 94, 699–711.

Goodman, J. (2008), "Who merits financial aid?: Massachusetts' Adams scholarship," *Journal of Public Economics*, 92, 2121–2131.

Hjalmarsson, R. (2009), "Juvenile jails: A path to the straight and narrow or to hardened criminality?" *Journal of Law and Economics*, 52, 779–809.

Hoekstra, M. (2009), "The effect of attending the flagship state university on earnings: A discontinuity-based approach," *The Review of Economics and Statistics*, 91, 717–724.

Hoxby, C. M. (2000), "The effects of class size on student achievement: New evidence from population variation," *Quarterly Journal of economics*, 115, 1239–1285.

Imbens, G. W., and Rubin, D. B. (2015), *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press.

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

Keele, L. J., and Titiunik, R. (2015), "Geographic Boundaries as Regression Discontinuities," *Political Analysis*, 23, 127–155.

Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016), "Field of study, earnings, and self-selection," *The Quarterly Journal of Economics*, 131, 1057–1111.

Klašnja, M., and Titiunik, R. (2017), "The incumbency curse: Weak parties, term limits, and unfulfilled accountability," *American Political Science Review*, 111, 129–148.

Litschig, S., and Morrison, K. M. (2013), "The impact of intergovernmental transfers on education outcomes and poverty reduction," *American Economic Journal: Applied Economics*, 5, 206–240.

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

Reardon, S. F., and Robinson, J. P. (2012), "Regression discontinuity designs with multiple rating-score variables," *Journal of Research on Educational Effectiveness*, 5, 83–104.

Rosenbaum, P. R. (2010), *Design of Observational Studies*, New York: Springer.

Snider, C., and Williams, J. W. (2015), "Barriers to entry in the airline industry: A multidimensional regression-discontinuity analysis of AIR-21," *Review of Economics and Statistics*, 97, 1002–1022.

Spenkuch, J. L., and Toniatti, D. (2018), "Political Advertising and Election Results," *The Quarterly Journal of Economics*.

Urquiola, M. (2006), "Identifying class size effects in developing countries: Evidence from rural Bolivia," *Review of Economics and Statistics*, 88, 171–177.

Urquiola, M., and Verhoogen, E. (2009), "Class-size caps, sorting, and the regression-discontinuity design," *The American Economic Review*, 99, 179–215.

Van der Klaauw, W. (2002), "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression–Discontinuity Approach*," *International Economic Review*, 43, 1249–1287.

——— (2008), "Breaking the link between poverty and low student achievement: An evaluation of Title I," *Journal of Econometrics*, 142, 731–756.