Supplemental Appendix to "Interpreting Regression Discontinuity Designs with Multiple Cutoffs"*

Matias D. Cattaneo† Luke Keele‡ Rocío Titiunik§ Gonzalo Vazquez-Bare¶

March 23, 2016

Abstract

This supplemental appendix contains a literature review on RD designs employing data with multiple cutoffs, the proofs of our main results, additional methodological results, and further empirical evidence not included in the main paper to conserve space.

*Cattaneo and Titiunik gratefully acknowledge financial support from the National Science Foundation through grant SES 1357561.

†Associate Professor, Department of Economics and Department of Statistics, University of Michigan.

‡Associate Professor, Department of Political Science, Penn State University.

§Corresponding author. Assistant Professor, Department of Political Science, University of Michigan.

¶Ph.D. candidate, Department of Economics, University of Michigan.

Contents

S1 Overview

This document includes additional material not included in the paper "Interpreting Regression Discontinuity Designs with Multiple Cutoffs" to conserve space.

Section S2 reports a selected list of RD papers employing data with multiple cutoffs in political science, economics and other disciplines, where the predominant strategy for identification, estimation and inference is the normalizing-and-pooling approach.

Section S3 provides the proofs of the results presented in the paper.

Finally, Section S4 gives some extensions and further discussion. In particular, Section S4.1 compares the different weighting schemes in the pooled estimand and the overall average of treatment effects across cutoffs, and contrasts the parameters with the ones in Lee (2008). Section S4.2 extends our results for sharp multi-cutoff RD designs to the case of kink multi-cutoff RD designs (c.f., Card, Lee, Pei, and Weber, 2015). Finally, section S4.3 discusses the relationship between RD designs with multiple cutoffs and multidimensional RD designs, i.e., RD designs with multiple running variables (c.f., Papay, Willett, and Murnane, 2011; Wong, Steiner, and Cook, 2013; Keele and Titiunik, 2015).

S2 Literature Review

Table S1 provides a selected list of examples of empirical papers employing RD designs with multiple cutoffs in Political Science and other disciplines, including economics, education, public health and public policy. In most cases, these papers apply only the normalization-and-pooling approach.

S3 Proofs of Results

This section gives the proofs and derivations underlying the main results reported in the paper. We employ the same notation and assumptions described in the paper, which are not reproduced here for brevity.

Table S1: Empirical Examples of Multi-Cutoff RD Designs with Normalization and Pooling

Citation	**Place**	**Score**	**Outcome**	**No. Cutoffs**
Political Science				
Albouy (2013)	U.S.	Vote Share	Federal Spending	Many
Boas and Hidalgo (2011)	Brazil	Vote Share	Incumbency	Many
Boas, Hidalgo, and Richardson (2014)	Brazil	Vote Share	Govt Contracts	Many
Brollo and Nannicini (2012)	Brazil	Vote Share	Federal Transfers	Many
Broockman (2009)	U.S.	Vote Share	Reverse Coattails	Many
Butler (2009)	U.S.	Vote Share	Incumbency	Many
Duraisamy, Lemennicier, and Khouri (2014)	India	Vote Share	Incumbency	Many
Eggers and Hainmueller (2009)	UK	Vote Share	Wealth	Many
Eggers et al. (2015)	Several	Vote Share	Incumbency	Many
Ferreira and Gyourko (2009)	U.S.	Vote Share	Policy Outcomes	Many
Folke and Snyder (2012)	U.S.	Vote Share	Gov. Vote Share	Many
Gagliarducci and Paserman (2012)	Italy	Vote Share	Early Termination	Many
Gerber and Hopkins (2011)	U.S.	Vote Share	Municipal Spending	Many
Hainmueller and Kern (2008)	Germany	Vote Share	Incumbency	Many
Kendall and Rekkas (2012)	Canada	Vote Share	Incumbency	Many
Klašnja (2015)	Romania	Vote Share	Incumbency	Many
Klašnja and Titiunik (2016)	Brazil	Vote Share	Incumbency	Many
Lee, Moretti, and Butler (2004)	U.S.	Vote Share	Incumbency	Many
Lee (2008)	U.S.	Vote Share	Incumbency	Many
Pettersson-Lidbom (2008)	Sweden	Vote Share	Fiscal Policy	Many
Trounstine (2011)	U.S.	Vote Share	Incumbency	Many
Uppal (2009)	India	Vote Share	Incumbency	Many
Uppal (2010)	U.S.	Vote Share	Incumbency	Many
Education				
Angrist and Lavy (1999)	Israel	Cohort size	Test scores	3
Canton and Blom (2004)	Mexico	Eligibility Score	College Outcomes	5
Chay, McEwan, and Urquiola (2005)	Chile	Eligibility Score	School Aid	13
Dobkin and Ferreira (2010)	U.S.	Birthday	Education Attainment	3
Goodman (2008)	U.S.	Test Score	Scholarship	20+
Hoxby (2000)	U.S.	Cohort size	Test scores	3
Kane (2003)	U.S.	GPA	College Attendance	4
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	3
Van der Klaauw (2008)	U.S.	Poverty Score	School Aid	3
Criminal Justice				
Berk and de Leeuw (1999)	U.S.	Prison Score	Re-conviction	4
Chen and Shapiro (2004)	U.S.	Prison Score	Rearrest	4
Hjalmarsson (2009)	U.S.	Adjudication Score	Re-conviction	2
Miscellaneous				
Behaghel, Crépon, and Sédillot (2008)	France	Age	Layoff Rates	20+
Black, Galdo, and Smith (2007)	U.S.	Training Eligibility	Job Training and Aid	2
Brollo et al. (2013)	Brazil	Population	Federal Transfers	7
Buddelmeyer and Skoufias (2004)	Mexico	Poverty Score	Education Attainment	7
Card and Shore-Sheppard (2004)	U.S.	Child Age and Income	Dr. Visits	4
Chen and Van der Klaauw (2008)	U.S.	Age	Disability Awards	3
Edmonds (2004)	S. Africa	Age	Child Outcomes	3
Litschig and Morrison (2013)	Brazil	Population	Poverty Reduction	17

Note: "Many" refers to examples based on vote shares, where the cutoff is a continuous random variable; in these cases, the number of cutoffs is related to the number of effective parties.

S3.1 Lemma 1: Pooled Sharp Multi-Cutoff RD

Fix $\varepsilon > 0$. Because the design is sharp, we have that

$$\begin{aligned}\mathbb{E}[Y_i \mid \tilde{X}_i = \varepsilon] &= \mathbb{E}\left\{\mathbb{E}[Y_i \mid X_i - C_i = \varepsilon, C_i] \mid \tilde{X}_i = \varepsilon\right\} \\ &= \sum_{c\in\mathcal{C}} \mathbb{E}[Y_i \mid X_i - C_i = \varepsilon, C_i = c]\mathbb{P}[C_i = c \mid \tilde{X}_i = \varepsilon] \\ &= \sum_{c\in\mathcal{C}} \mathbb{E}[Y_{1i}(c) \mid X_i = c + \varepsilon, C_i = c]\mathbb{P}[C_i = c \mid \tilde{X}_i = \varepsilon]\end{aligned}$$

and similarly

$$\mathbb{E}[Y_i \mid \tilde{X}_i = -\varepsilon] = \sum_{c\in\mathcal{C}} \mathbb{E}[Y_{0i}(c) \mid X_i = c - \varepsilon, C_i = c]\mathbb{P}[C_i = c \mid \tilde{X}_i = -\varepsilon]$$

On the other hand,

$$\mathbb{P}[C_i = c \mid \tilde{X}_i = x] = \frac{f_{\tilde{X}|C}(x|c)\mathbb{P}[C_i = c]}{f_{\tilde{X}}(x)} = \frac{f_{X|C}(c + x|c)\mathbb{P}[C_i = c]}{\sum_{c\in\mathcal{C}} f_{X|C}(c + x|c)\mathbb{P}[C_i = c]}$$

Define $\Delta(\varepsilon) = \mathbb{E}[Y_i|\tilde{X}_i = \varepsilon] - \mathbb{E}[Y_i|\tilde{X}_i = -\varepsilon]$. Since the support of C_i is finite, interchanging limits and sums is allowed. Hence, by continuity of the conditional expectation functions and densities, taking limit as $\varepsilon \to 0^+$ leads to

$$\tau^{\mathsf{P}} = \lim_{\varepsilon\to 0^+} \Delta(\varepsilon) = \sum_{c\in\mathcal{C}} \mathbb{E}[Y_{1i}(c) - Y_{0i}(c) \mid X_i = c, C_i = c] \cdot \frac{f_{X|C}(c|c)\mathbb{P}[C_i = c]}{\sum_{c\in\mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C_i = c]}$$

Define $\omega(c) := \frac{f_{X|C}(c|c)\mathbb{P}[C_i=c]}{\sum_{c\in\mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C_i=c]}$ and the result follows.

S3.2 Proposition 1: Constant Treatment Effects

Under the assumption of constant treatment effects within cutoffs, $Y_{1i}(c) - Y_{0i}(c) = \tau(c)$, and continuity of the conditional expectations holds automatically. Hence, $\mathbb{E}[Y_{1i}(c) - Y_{0i}(c) \mid X_i = c, C_i = c] = \tau(c)$ and the result follows from Lemma 1.

S3.3 Proposition 2: Score-Ignorable Treatment Effects

Under score Ignorability, $\mathbb{E}[Y_{1i}(c) - Y_{0i}(c) \mid X_i, C_i = c] = \mathbb{E}[Y_{1i}(c) - Y_{0i}(c) \mid C_i = c]$, and continuity of the conditional expectations holds automatically. The result follows from Lemma 1.

S3.4 Proposition 3: Cutoff-Ignorable Treatment Effects

Under cutoffs Ignorability, $\mathbb{E}[Y_{1i}(c)-Y_{0i}(c) \mid X_i, C_i = c] = \mathbb{E}[Y_{1i}-Y_{0i} \mid X_i = c]$ and the result follows from Lemma 1.

S3.5 Lemma 2: Pooled Multi-Cutoff Fuzzy RD

Fix $\varepsilon > 0$. Taking the first term in the numerator,

$$\begin{aligned}\mathbb{E}[Y_i \mid \tilde{X}_i = \varepsilon] &= \mathbb{E}\left[\mathbb{E}[Y_i \mid X_i - C_i = \varepsilon, C_i] \mid \tilde{X}_i = \varepsilon\right] \\ &= \sum_{c\in\mathcal{C}} \mathbb{E}[Y_i \mid X_i - C_i = \varepsilon, C_i = c]\mathbb{P}[C_i = c \mid \tilde{X}_i = \varepsilon]\end{aligned}$$

Now, we have that

$$\begin{aligned}&\mathbb{E}[Y_i \mid X_i - C_i = \varepsilon, C_i = c] \\ &\quad = \mathbb{E}[Y_i \mid X_i = c+\varepsilon, C_i = c] \\ &\quad = \mathbb{E}[(Y_{1i}(c) - Y_{0i}(c))D_i \mid X_i = c+\varepsilon, C_i = c] + \mathbb{E}[Y_{0i}(c) \mid X_i = c+\varepsilon, C_i = c] \\ &\quad = \mathbb{E}[(Y_{1i}(c) - Y_{0i}(c))D_{1i}(c+\varepsilon, c) \mid X_i = c+\varepsilon, C_i = c] \\ &\qquad + \mathbb{E}[Y_{0i}(c) \mid X_i = c+\varepsilon, C_i = c]\end{aligned}$$

and so, by right continuity,

$$\begin{aligned}\Delta^+(c) \equiv \lim_{\varepsilon\to 0^+} \mathbb{E}[Y_i \mid X_i - C_i = \varepsilon, C_i = c] &= \mathbb{E}[(Y_{1i}(c) - Y_{0i}(c))D_{1i}(c) \mid X_i = c, C_i = c] \\ &\quad + \mathbb{E}[Y_{0i}(c) \mid X_i = c, C_i = c]\end{aligned}$$

By an analogous reasoning,

$$\begin{aligned}\mathbb{E}[Y_i \mid X_i - C_i = -\varepsilon, C_i = c] &= \mathbb{E}[(Y_{1i}(c) - Y_{0i}(c))D_{0i}(c-\varepsilon, c) \mid X_i = c-\varepsilon, C_i = c] \\ &\quad + \mathbb{E}[Y_{0i}(c) \mid X_i = c-\varepsilon, C_i = c]\end{aligned}$$

and hence by left continuity,

$$\Delta^-(c) \equiv \lim_{\varepsilon\to 0^+} \mathbb{E}[Y_i \mid X_i - C_i = -\varepsilon, C_i = c] = \mathbb{E}[(Y_{1i}(c) - Y_{0i}(c))D_{0i}(c) \mid X_i = c, C_i = c]$$
$$+ \mathbb{E}[Y_{0i}(c) \mid X_i = c, C_i = c]$$

giving

$$\begin{aligned}\Delta^+(c) - \Delta^-(c) &= \mathbb{E}[(Y_{1i}(c) - Y_{0i}(c))(D_{1i} - D_{0i}) \mid X_i = c, C_i = c] \\ &= \mathbb{E}[(Y_{1i}(c) - Y_{0i}(c)) \mid D_{1i} > D_{0i}, X_i = c, C_i = c]\mathbb{P}[D_{1i} > D_{0i} \mid X_i = c, C_i = c]\end{aligned}$$

where the second equality follows by monotonicity. On the other hand, by previous calculations,

$$\mathbb{P}[C_i = c \mid \tilde{X}_i = x] = \frac{f_{X|C}(c + x|c)\mathbb{P}[C_i = c]}{\sum_{c\in\mathcal{C}} f_{X|C}(c + x|c)\mathbb{P}[C_i = c]}$$

and by continuity,

$$\lim_{\varepsilon\to 0^+} \mathbb{P}[C_i = c \mid \tilde{X}_i = \varepsilon] = \frac{f_{X|C}(c|c)\mathbb{P}[C_i = c]}{\sum_{c\in\mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C_i = c]}$$

For the denominator we have that:

$$\mathbb{E}[D_i \mid \tilde{X}_i = \varepsilon] = \sum_{c\in\mathcal{C}} \mathbb{E}[D_i \mid X_i - C_i = \varepsilon, C_i = c]\mathbb{P}[C_i = c \mid \tilde{X}_i = \varepsilon]$$

But

$$\mathbb{E}[D_i \mid X_i - C_i = \varepsilon, C_i = c] = \mathbb{E}[D_{1i}(c + \varepsilon, c) \mid X_i = c + \varepsilon, C_i = c]$$

so by continuity,

$$D^+(c) \equiv \lim_{\varepsilon\to 0^+} \mathbb{E}[D_i \mid X_i - C_i = \varepsilon, C_i = c] = \mathbb{E}[D_{1i}(c) \mid X_i = c, C_i = c]$$

and similarly

$$D^-(c) \equiv \lim_{\varepsilon\to 0^+} \mathbb{E}[D_i \mid X_i - C_i = -\varepsilon, C_i = c] = \mathbb{E}[D_{0i}(c) \mid X_i = c, C_i = c]$$

which gives

$$D^+(c) - D^-(c) = \mathbb{E}[D_{1i}(c) - D_{0i}(c) \mid X_i = c, C_i = c] = \mathbb{P}[D_{1i}(c) > D_{0i}(c) \mid X_i = c, C_i = c]$$

Combining all the terms,

$$\tau^{\mathtt{P}}_{\mathtt{FRD}} = \sum_{c \in \mathcal{C}} \mathbb{E}[Y_{1i}(c)(c) - Y_{0i}(c)(c) \mid D_{1i}(c) > D_{0i}(c), X_i = c, C_i = c]\; \omega_{\mathtt{F}}(c)$$

where

$$\omega_{\mathtt{FRD}}(c) = \frac{\mathbb{P}[D_{1i}(c) > D_{0i}(c) \mid X_i = c, C_i = c] f_{X|C}(c|c)\mathbb{P}[C_i = c]}{\sum\limits_{c \in \mathcal{C}} \mathbb{P}[D_{1i}(c) > D_{0i}(c) \mid X_i = c, C_i = c] f_{X|C}(c|c)\mathbb{P}[C_i = c]}$$

which completes the proof.

The continuity Assumption 7 may be hard to interpret as it involves a random variable that is a combination of potential outcomes and potential treatment statuses. A stronger but more easily interpretable condition is the following:

- $\mathbb{E}[Y_{di}(c) \mid D_{0i}(c) = d_0, D_{1i}(c) = d_1, X_i = x, C_i = c]$ and $\mathbb{P}[D_{0i}(c) = d_0, D_{1i}(c) = d_1 \mid X_i = x, C_i = c]$ are continuous in x at $x = c$ for $d, d_0, d_1 \in \{0, 1\}$.

S4 Extensions and Further Discussion

S4.1 Pooled Estimand versus Average of Cutoff-Specific Effects

To further understand $\tau^{\mathtt{P}}$, it is useful to contrast it with the overall average of the (average) treatment effects at every cutoff. This overall average of all the cutoff-specific effects is given by

$$\bar{\tau} = \sum_{c \in \mathcal{C}} \mathbb{E}[Y_{1i}(c) - Y_{0i}(c) \mid X_i = c, C_i = c]\mathbb{P}[C_i = c]$$

These two effects are different due to the presence of $f_{X|C}(c|c)$ in the pooled estimand. In $\bar{\tau}$, the weights are simply the probability that the random cutoff C takes each particular value c. In contrast, in $\tau^{\mathtt{P}}$, this probability is multiplied by the factor $\frac{f_{X|C}(c|c)}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C_i = c]}$, which depends on $f_{X|C}(c|c)$, the conditional density of the running variable given C. Suppose the potential outcomes can be written as nonseparable functions:

$$Y_{1i}(c) = y_1(x, c, U_i), \quad Y_{0i}(c) = y_0(x, c, U_i)$$

where the variable U_i captures individual heterogeneity or the "type" of the individual. Define:

$$y_1^+(c, u) \equiv \lim_{x \to c^+} y_1(x, c, u), \quad y_1^-(c, u) \equiv \lim_{x \to c^-} y_1(x, c, u)$$

Then we can write the average treatment effect at each cutoff as:

$$\begin{aligned}
\mathbb{E}[Y_{1i}(c) - Y_{0i}(c) \mid X_i = c, C_i = c] &= \int (y_1^+(c, u) - y_0^-(c, u))\, \mathrm{d}F_{U|X,C}(u|c, c) \\
&= \int (y_1^+(c, u) - y_0^-(c, u)) \frac{f_{X,C|U}(c, c|u)}{f_{X,C}(c, c)}\, \mathrm{d}F_U(u) \\
&= \int (y_1^+(c, u) - y_0^-(c, u)) \frac{f_{X|C,U}(c|c, u)}{f_{X|C}(c|c)} \cdot \frac{\mathbb{P}[C = c|u]}{\mathbb{P}[C = c]}\, \mathrm{d}F_U(u)
\end{aligned}$$

where $\mathbb{P}[C = c|u]$ is the probability that $C = c$ conditional on $U = u$ and $f_{X|C,U}(c|c, u)$ is the density of X conditional on $C = c$ and $U = u$. This treatment effect is calculated as a weighted average of individual effects at $X = C$ for the whole population (not only for units around the cutoff), where the weights are higher for units who are more likely to face that particular cutoff and for units who, conditional on facing the cutoff, are more likely to be around the threshold. In particular, if C is independent of both X and U and the exclusion restriction holds, this parameter becomes the one in Lee (2008).

In this setting, our results show that the pooled estimand can be written as:

$$\tau^{\mathsf{P}} = \sum_{c \in \mathcal{C}} \int (y_1^+(c, u) - y_0^-(c, u)) \frac{f_{X|C,U}(c|c, u)}{f_{X|C}(c|c)} \cdot \frac{f_{X|C}(c|c)}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C = c]} \cdot \mathbb{P}[C = c|u]\, \mathrm{d}F_U(u)$$

On the other hand, the average of (average) treatment effects over cutoffs is:

$$\bar{\tau} = \int (y_1^+(c, u) - y_0^-(c, u)) \frac{f_{X|C,U}(c|c, u)}{f_{X|C}(c|c)} \cdot \mathbb{P}[C = c|u]\, \mathrm{d}F_U(u)$$

Thus, the pooled estimand differs from the average over cutoffs by the term $\frac{f_{X|C}(c|c)}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C=c]}$, which is the density of the running variable at each cutoff relative to the average conditional density. Compared to $\bar{\tau}$, the pooled estimand gives more weight to the effects at the values c for which this density is above its average $\sum_{c\in\mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C = c]$. If this conditional density is constant over cutoffs, i. e. $f_{X|C}(c|c) = k$ for all $c \in \mathcal{C}$, where k is a constant, this additional weighting factor becomes one, and $\tau^{\mathsf{P}} = \bar{\tau}$.

S4.2 Kink RD Design with Multiple Cutoffs

In this section, we succinctly show how our results can be extended to the Kink RD design. This design arises when a treatment or policy is assigned on the basis of a score via a formula, and this formula contains one or more kinks—points at which the formula that relates the assignment variable to the treatment changes. For example, unemployment insurance benefits may be 100 dollars for individuals with one dependent, 200 dollars for individuals with two dependents, and 300 dollars for individuals with three or more dependents, creating a piece-wise linear relationship between number of dependents and benefits.

Formally, these kinks in the formula that connects the assignment variable (number of dependents) to the treatment (unemployment insurance benefits) are discontinuous jumps in the first derivative or slope of the conditional regression function of the treatment given the assignment variable at the points in the assignment variable where the kinks occur. The kink RD design is analyzed formally by Card, Lee, Pei, and Weber (2015), who discuss nonparametric identification results. To our knowledge, kink RD design applications have not yet been explored in political science, but we imagine that kinks in policy rules could be exploited, for example, to study whether increased welfare benefits translate into increased political support of the party who is seen to "own" that particular issue area.

Let the outcome variable be $Y = y(B, X, C, U)$ where B is a (continuous) treatment of interest such as unemployment benefits and, as before, X, C and U represent the running variable, the cutoff that each individual faces and the individual heterogeneity, respectively. For the moment we focus on the case of discrete cutoffs. The treatment of interest is now a function of two arguments, X and C: $B = b(X, C)$.

We start by assuming that we have a multi-cutoff RKD design, that is, that there is a discontinuity in the first derivative of the regression function at each possible value of the cutoff:

Assumption 1 (Kink RD) *For all $c \in \mathcal{C}$:*

$\lim_{x \to c^+} \frac{\partial}{\partial x} b(x, c) \neq \lim_{x \to c^-} \frac{\partial}{\partial x} b(x, c)$.

As before, let $\tilde{X} = X - C$. A pooling approach would sum all individuals with the same value of the running variable across cutoffs, i.e., to use as treatment of interest the variable $b(x) = \sum_{c \in \mathcal{C}} b(x, c)$, which in turn implies $\frac{d}{dx} b(x) = \sum_{c \in \mathcal{C}} \frac{\partial}{\partial x} b(x, c)$. We can define the pooled estimand as:

$$\tau^{\mathtt{P}}_{\mathtt{KRD}} = \frac{\lim\limits_{x \to 0^+} \frac{d}{dx} \mathbb{E}[Y \mid \tilde{X} = x] - \lim\limits_{x \to 0^-} \frac{d}{dx} \mathbb{E}[Y \mid \tilde{X} = x]}{\lim\limits_{x \to 0^+} \frac{d}{dx} b(x) - \lim\limits_{x \to 0^-} \frac{d}{dx} b(x)}.$$

This corresponds to the sharp case, and the fuzzy case can be analyzed analogously.

We denote by $y_1(b, x, c)$ and $y_2(b, x, c)$ the derivatives of $y(b, x, c)$ with respect to its first and second arguments, respectively. We summarize the results for the multi-cutoff RK design in the following lemma:

Lemma 1 (Kink Multi-Cutoff RD) *Suppose the following assumptions hold:*

1. *$y(b, x, c)$ is continuous in b and x, with $y_1(b, x, c) \equiv \frac{\partial y}{\partial b}$ continuous in b*
2. *$y_2(b, x, c) \equiv \frac{\partial y}{\partial x}$ is continuous in x, $\forall b$*
3. *$b(x, c)$ is a known function that is continuously differentiable with respect to x, except at $x = c$, where $\forall c \in \mathcal{C}$, $\lim_{x \to c^+} \frac{\partial}{\partial x} b(x, c) \neq \lim_{x \to c^-} \frac{\partial}{\partial x} b(x, c)$.*
4. *The density of X conditional on $C = c$ and $U = u$, $f_{X|C,U}(x|c, u)$, is continuously differentiable with respect to x for all c and u.*

Then, the pooled kink RD causal estimand is

$$\tau^{\mathtt{P}}_{\mathtt{KRD}} = \sum_{c \in \mathcal{C}} \mathbb{E}[y_{1i}(b^c_0, 0, c) \mid X = c, C = c] \, \omega_{\mathtt{KRD}}(c)$$

where

$$\omega_{\mathtt{KRD}}(c) = \frac{\lim_{x \to 0^+} \frac{\partial}{\partial x} b(x, c) - \lim_{x \to 0^-} \frac{\partial}{\partial x} b(x, c)}{\sum_{c \in \mathcal{C}} \left(\lim_{x \to 0^+} \frac{\partial}{\partial x} b(x, c) - \lim_{x \to 0^-} \frac{\partial}{\partial x} b(x, c) \right)} \cdot \frac{f_{X|C}(c|c) \mathbb{P}[C = c]}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c) \mathbb{P}[C = c]}$$

The proof of this lemma is as follows. Using the product rule, the first term in the numerator becomes

$$\frac{d}{dx}\mathbb{E}[Y \mid \tilde{X}=x] = \frac{d}{dx}\sum_{c\in\mathcal{C}}\mathbb{E}[Y \mid X=x+c, C=c]\mathbb{P}[C=c \mid X-C=x]$$
$$= \sum_{c\in\mathcal{C}}\frac{\partial}{\partial x}b(x,c)\mathbb{E}[y_1(b,x,c,w) \mid X=x+c, C=c]\mathbb{P}[C=c \mid X-C=x]$$
$$+ \sum_{c\in\mathcal{C}}\mathbb{E}[y_2(b,x,c,w) \mid X=x+c, C=c]\mathbb{P}[C=c \mid X-C=x]\mathbb{P}[C=c \mid X-C=x]$$
$$+ \sum_{c\in\mathcal{C}}\int y(b,x,c,w)\frac{\partial}{\partial x}f_{U|X,C}(w \mid x,c)\mathrm{d}w$$
$$+ \sum_{c\in\mathcal{C}}\mathbb{E}[Y \mid \tilde{X}=x]\frac{d}{dx}\mathbb{P}[C=c \mid X-C=x]$$

and similarly for the second term. Under the continuity assumptions above, all the terms except for the first one cancel out when taking the difference, which yields

$$\lim_{x\to 0^+}\frac{d}{dx}\mathbb{E}[Y \mid \tilde{X}=x] - \lim_{x\to 0^-}\frac{d}{dx}\mathbb{E}[Y \mid \tilde{X}=x]$$
$$= \sum_{c\in\mathcal{C}}\left(\lim_{x\to 0^+}\frac{\partial}{\partial x}b(x,c) - \lim_{x\to 0^-}\frac{\partial}{\partial x}b(x,c)\right)\mathbb{E}[y_1(b_0^c,0,c,w)]\mathbb{P}[C=c \mid X=c]$$

where $b_0^c = b(0,c)$. Finally, the denominator is simply

$$\lim_{x\to 0^+}\frac{d}{dx}b(x) - \lim_{x\to 0^-}\frac{d}{dx}b(x) = \sum_{c\in\mathcal{C}}\left(\lim_{x\to 0^+}\frac{\partial}{\partial x}b(x,c) - \lim_{x\to 0^-}\frac{\partial}{\partial x}b(x,c)\right).$$

which gives the desired result.

S4.3 Comparison with Multidimensional Scores

We now briefly describe the connections between RD designs with multiple scores or running variables and the multi-cutoff RD design that we explore in this paper. For concreteness, we focus on the Geographic RD design, where there are two adjacent geographic areas separated by a boundary and the treatment is assigned to all units in one area and withheld from all units in the other. The GRD design is discussed in Keele and Titiunik (2015) and a discussion of RD designs with multiple running variables can be found in (Papay, Willett, and Murnane, 2011; Wong, Steiner, and Cook, 2013).

As before, our point of departure is the pooled parameter in the sharp design, $\tau^{\mathsf{P}} = \sum_{c\in\mathcal{C}}\mathbb{E}[Y_{1i}(c)-$

$Y_{0i}(c) \mid X_i = c, C_i = c]\omega(c)$ where, again, the weights are $\omega(c) = \frac{f_{X|C}(c|c)\mathbb{P}[C_i=c]}{\sum_{c\in\mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C_i=c]}$. These weights can be rewritten as:

$$\omega(c) = \frac{f_{XC}(c,c)}{\sum_{c\in\mathcal{C}} f_{XC}(c,c)}$$

where $f_{XC}(x,c)$ is the (mixed) joint density of (X,C). Now call $\mathbf{S}_i = (X_i, C_i)$ the vector containing the running variable and the cutoff, and $\mathbf{c} = (c,c)$ the value of $\mathbf{S}_i$ when $X_i = C_i = c$. Finally, denote the treatment effect at each cutoff by $\tau(\mathbf{c}) = \mathbb{E}[Y_{1i}(c) - Y_{0i}(c) \mid X_i = c, C_i = c] = \mathbb{E}[Y_{1i}(c) - Y_{0i}(c) \mid \mathbf{S}_i = \mathbf{c}]$. Then, we have

$$\tau^{\mathsf{P}} = \sum_{c\in\mathcal{C}} \tau(\mathbf{c}) \frac{f_{XC}(\mathbf{c})}{\sum_{c\in\mathcal{C}} f_{XC}(\mathbf{c})} = \frac{\sum_{c\in\mathcal{C}} \tau(\mathbf{c}) f_S(\mathbf{c})}{\sum_{c\in\mathcal{C}} f_S(\mathbf{c})}.$$

Now, if we define: $\mathcal{B} = \{(X,C) : X = c, C = c\}$ we get

$$\tau^{\mathsf{P}} = \frac{\sum_{\mathbf{s}\in\mathcal{B}} \tau(\mathbf{s}) f_S(\mathbf{s})}{\sum_{\mathbf{s}\in\mathcal{B}} f_S(\mathbf{s})}$$

This expression is the discrete analog of the expressions in Papay, Willett, and Murnane (2011), Wong, Steiner, and Cook (2013), and Keele and Titiunik (2015), and it shows that we can interpret the multi-cutoff RD design as an RD design with two running variables X and C where the boundary is the set of points at which $X = C$. In other words, an RD with one score and multiple cutoffs can be recast as an RD with two running variables.

S5 Comparison of U.S. Senate Elections to Brazilian Mayoral Elections

We present a descriptive analysis to show that exploiting the multi-cutoff structure of the election-based RD design is possible in the example based on Brazilian mayoral elections but not in the example based on U.S. Senate elections. In the latter, there are simply not enough observations for lower values of the cutoff variable.

As we highlighted in the main paper, these two examples differ sharply in the density of observations at different cutoff values. There are very few U.S. Senate elections where a third party obtains anything more than a very small fraction of the vote. In the Brazilian mayoral elections, however, about a third of races occurs in municipalities where the two top-getters combined obtain less than 70% of the vote. Table S2 presents the frequency of races in our sample by different levels of strongest opponent's vote shares at t for the Democratic Party and the PSDB. Since this variable is continuous, we divide its support in four exclusive intervals: $[0, 35)$, $[35, 40)$, $[40, 45)$, and $[45, 50)$. Within each of these intervals of strongest opponent's vote share at t, Table S2 reports the number of elections that each party won and lost at t. Note that in a perfect two-party system, knowing the value of a party's strongest opponent's vote share is *equivalent* to knowing whether the party won or lost the election, but this equivalency is broken in a multi-party RD design.

For example, the columns corresponding to the PSDB show that, of the 1346 races in our sample where the PSBD's strongest opponent obtained between 35% and 40% of the t vote, the PSDB won roughly 85% and lost the rest. The proportion of victories decreases for higher values of this variable, with the PSDB winning no more than 64% of the races in all cells where vote share of its strongest opponent is 35% or higher.

Table S2: Frequency of Observations for Different Levels of Strongest Opponent's Vote Shares at t

	Democratic Party U.S. Senate Elections			PSDB Brazil Mayoral Elections		
Opponent Vote (%)	Total	Victories (%)	Defeats (%)	Total	Victories (%)	Defeats (%)
$[0, 35)$	264	100.0	0.0	1346	84.9	15.1
$[35, 40)$	118	94.1	5.9	986	63.9	36.1
$[40, 45)$	161	96.3	3.7	1251	62.3	37.7
$[45, 50)$	221	77.8	22.2	1490	61.5	38.5

Note: Columns corresponding to Democratic Party report number of U.S. Senate elections in 1910–2010. Columns corresponding to PSDB report number of mayoral elections in Brazil in 1996-2012.

A very different situation occurs in U.S. Senate elections where, for example, the Democratic Party won all 264 races where the strongest opponent obtained less than 35% of the vote, as would occur in a perfect two-party system. Similarly, of the 118 races in our sample where the Democratic Party's strongest opponent obtained between 35% and 40% of the t vote, the party won 111 and lost only 7. It is only in the $[45, 50)$ range where the party loses 20% of races—a non-neligible but still small proportion. Thus, despite third candidates being common, RD designs based on U.S. Senate elections behave as single-cutoff because most races are decided very close to the 50% cutoff.

References

Albouy, David. 2013. "Partisan Representation in Congress and the Geographic Distribution of Federal Funds." *Review of Economics and Statistics* 95 (1):127–141.

Angrist, Joshua D and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114 (2):533–575.

Behaghel, Luc, Bruno Crépon, and Béatrice Sédillot. 2008. "The Perverse Effects of Partial Employment Protection Reform: The Case of French Older Workers." *Journal of Public Economics* 92 (3):696–721.

Berk, Richard A and Jan de Leeuw. 1999. "An Evaluation of California's Inmate Classification System Using a Generalized Regression Discontinuity Design." *Journal of the American Statistical Association* 94 (448):1045–1052.

Black, Dan A, Jose Galdo, and Jeffrey A Smith. 2007. "Evaluating the Worker Profiling and Reemployment Services System Using a Regression Discontinuity Approach." *American Economic Review* 97 (2):104–107.

Boas, Taylor C and F Daniel Hidalgo. 2011. "Controlling the Airwaves: Incumbency Advantage and Community Radio in Brazil." *American Journal of Political Science* 55 (4):869–885.

Boas, Taylor C, F Daniel Hidalgo, and Neal P Richardson. 2014. "The Spoils of Victory: Campaign Donations and Government Contracts in Brazil." *Journal of Politics* 76 (2):415–429.

Brollo, Fernanda and Tommaso Nannicini. 2012. "Tying Your Enemy's Hands in Close Races: The Politics of Federal Transfers in Brazil." *American Political Science Review* 106 (40):742–761.

Brollo, Fernanda, Tommaso Nannicini, Roberto Perotti, and Guido Tabellini. 2013. "The Political Resource Curse." *American Economic Review* 103 (5):1759–96.

Broockman, David E. 2009. "Do Congressional Candidates Have Reverse Coattails? Evidence from a Regression Discontinuity Design." *Political Analysis* 17 (4):418–434.

Buddelmeyer, Hielke and Emmanuel Skoufias. 2004. *An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA*, vol. 827. World Bank Publications.

Butler, Daniel Mark. 2009. "A Regression Discontinuity Design Analysis of the Incumbency Advantage and Tenure in the US House." *Electoral Studies* 28 (1):123–128.

Canton, Erik and Andreas Blom. 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, David, David S. Lee, Zhuan Pei, and Andrea Weber. 2015. "Inference on Causal Effects in a Generalized Regression Kink Design." *Econometrica* 83 (6):2453–2483.

Card, David and Lara D Shore-Sheppard. 2004. "Using Discontinuous Eligibility Rules to Identify the Effects of the Federal Medicaid Expansions on Low-Income Children." *Review of Economics and Statistics* 86 (3):752–766.

Chay, Kenneth Y, Patrick J McEwan, and Miguel Urquiola. 2005. "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools." *American Economic Review* 95 (4):1237–1258.

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

Chen, Susan and Wilbert Van der Klaauw. 2008. "The Work Disincentive Effects of the Disability Insurance Program in the 1990s." *Journal of Econometrics* 142 (2):757–784.

Dobkin, Carlos and Fernando Ferreira. 2010. "Do School Entry Laws Affect Educational Attainment and Labor Market Outcomes?" *Economics of Education Review* 29 (1):40–54.

Duraisamy, P., Bertrand Lemennicier, and Michele Khouri. 2014. "The Incumbency Effect In The Indian Parliamentary Elections, 2004 and 2009: A Regression Discontinuity Approach." *Journal of Quantitative Economics* 12 (1):12–30.

Edmonds, Eric 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.

Eggers, Andrew, Anthony Fowler, Jens Hainmueller, Andrew B Hall, and James M Snyder. 2015. "On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races." *American Journal of Political Science* 59 (1):259–274.

Eggers, Andrew C. and Jens Hainmueller. 2009. "MPs for Sale? Returns to Office in Postwar British Politics." *American Political Science Review* 103 (4):513–533.

Ferreira, Fernando and Joseph Gyourko. 2009. "Do Political Parties Matter? Evidence from US Cities." *Quarterly Journal of Economics* 124 (1):399–422.

Folke, Olle and James M Snyder. 2012. "Gubernatorial Midterm Slumps." *American Journal of Political Science* 56 (4):931–948.

Gagliarducci, Stefano and M. Daniele Paserman. 2012. "Gender Interactions within Hierarchies: Evidence from the Political Arena." *Review of Economic Studies* 79 (3):1021–1052.

Gerber, Elisabeth R and Daniel J Hopkins. 2011. "When Mayors Matter: Estimating the Impact of Mayoral Partisanship on City Policy." *American Journal of Political Science* 55 (2):326–339.

Goodman, Joshua. 2008. "Who Merits Financial Aid?: Massachusetts' Adams Scholarship." *Journal of Public Economics* 92 (10):2121–2131.

Hainmueller, Jens and Holger Lutz Kern. 2008. "Incumbency as a Source of Spillover Effects in Mxed Electoral Systems: Evidence from a Regression-Discontinuity Design." *Electoral Studies* 27 (2):213–227.

Hjalmarsson, Randi. 2009. "Juvenile Jails: A Path to the Straight and Narrow or to Hardened Criminality?" *Journal of Law and Economics* 52 (4):779–809.

Hoxby, Caroline M. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics* 115 (4):1239–1285.

Kane, Thomas J. 2003. "A Quasi-Experimental Estimate of the Impact of Financial Aid on College-Going." Tech. rep., National Bureau of Economic Research.

Keele, Luke J. and Rocío Titiunik. 2015. "Geographic Boundaries as Regression Discontinuities." *Political Analysis* 23 (1):127–155.

Kendall, Chad and Marie Rekkas. 2012. "Incumbency Advantages in the Canadian Parliament." *Canadian Journal of Economics/Revue canadienne d'économique* 45 (4):1560–1585.

Klašnja, Marko. 2015. "Corruption and the Incumbency Disadvantage: Theory and Evidence." *Journal of Politics* 77 (4):928–942.

Klašnja, Marko and Rocío Titiunik. 2016. "The Incumbency Curse: Weak Parties, Term Limits, and Unfulfilled Accountability." *American Political Science Review*, forthcoming .

Lee, David S. 2008. "Randomized Experiments from Non-random Selection in U.S. House Elections." *Journal of Econometrics* 142 (2):675–697.

Lee, David S, Enrico Moretti, and Matthew J Butler. 2004. "Do Voters Affect or Elect Policies? Evidence from the US House." *Quarterly Journal of Economics* :807–859.

Litschig, Stephan and Kevin M Morrison. 2013. "The Impact of Intergovernmental Transfers on Education Outcomes and Poverty Reduction." *American Economic Journal: Applied Economics* 5 (4):206–240.

Papay, John P, John B Willett, and Richard J Murnane. 2011. "Extending the regression-discontinuity approach to multiple assignment variables." *Journal of Econometrics* 161 (2):203–207.

Pettersson-Lidbom, Per. 2008. "Do Parties Matter for Economic Outcomes? A Regression-Discontinuity Approach." *Journal of the European Economic Association* 6 (5):1037–1056.

Trounstine, Jessica. 2011. "Evidence of a Local Incumbency Advantage." *Legislative Studies Quarterly* 36 (2):255–280.

Uppal, Yogesh. 2009. "The Disadvantaged Incumbents: Estimating Incumbency Effects in Indian State Legislatures." *Public Choice* 138 (1-2):9–27.

———. 2010. "Estimating Incumbency Effects in U.S. State Legislatures: A Quasi-Experimental Study." *Economics & Politics* 22 (2):180–199.

Urquiola, Miguel. 2006. "Identifying Class Size Effects in Developing Countries: Evidence from Rural Bolivia." *Review of Economics and Statistics* 88 (1):171–177.

Urquiola, Miguel and Eric Verhoogen. 2009. "Class-size Caps, Sorting, and the Regression-Discontinuity Design." *American Economic Review* 99 (1):179–215.

Van der Klaauw, Wilbert. 2002. "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression–Discontinuity Approach." *International Economic Review* 43 (4):1249–1287.

———. 2008. "Breaking the Link between Poverty and Low Student Achievement: An Evaluation of Title I." *Journal of Econometrics* 142 (2):731–756.

Wong, Vivian C., Peter M. Steiner, and Thomas D. Cook. 2013. "Analyzing Regression-Discontinuity Designs With Multiple Assignment Variables A Comparative Study of Four Estimation Methods." *Journal of Educational and Behavioral Statistics* 38 (2):107–141.