Interpreting Regression Discontinuity Designs with Multiple Cutoffs

Matias D. Cattaneo, University of Michigan
Luke Keele, Penn State University
Rocío Titiunik, University of Michigan
Gonzalo Vazquez-Bare, University of Michigan

We consider a regression discontinuity (RD) design where the treatment is received if a score is above a cutoff, but the cutoff may vary for each unit in the sample instead of being equal for all units. This multi-cutoff regression discontinuity design is very common in empirical work, and researchers often normalize the score variable and use the zero cutoff on the normalized score for all observations to estimate a pooled RD treatment effect. We formally derive the form that this pooled parameter takes and discuss its interpretation under different assumptions. We show that this normalizing-and-pooling strategy so commonly employed in practice may not fully exploit all the information available in a multi-cutoff RD setup. We illustrate our methodological results with three empirical examples based on vote shares, population, and test scores.

The regression discontinuity (RD) design has become one of the preferred quasi-experimental research designs in the social sciences, mostly as a result of the relatively weak assumptions that it requires to recover causal effects. In the "sharp" version of the RD design, every subject is assigned a score and a treatment is given to all units whose score is above the cutoff and withheld from all units whose score is below it. Under the assumption that all possible confounders vary smoothly at the cutoff as a function of the score (also known as "running variable"), a comparison of units barely above and barely below the cutoff can be used to recover the causal effect of the treatment (for a review, see Skovron and Titiunik [2016] and references therein).

The RD design is widely used in political science. RD designs based on elections are particularly common, since the discontinuous assignment of victory in close races often provides a credible research design to make causal inferences about mass or elite behavior. Although the RD design has been found to fail in US House elections (Caughey and

Sekhon 2011), RD designs based on elections seem to be generally valid as an identification strategy to recover causal effects in other electoral contexts (Eggers et al. 2015; see de la Cuesta and Imai [2016] for further discussion). In addition to elections, RD designs in political science, as well as in other social and behavioral sciences, are based on other running variables such as population, test scores, poverty indexes, birth weight, geolocation, and income. For a list of examples of recent RD applications see section S2 in the appendix, available online.

In a standard RD design, the cutoff in the score that determines treatment assignment is known and equal for all units. For example, in the classic education example where a scholarship is awarded to students who score above a threshold on a standardized test, the cutoff for the scholarship is known and the same for every student. However, in many applications of the RD design, the value of the cutoff may vary by unit. One of the most common examples of variable cutoffs occurs in political science applications where the score is a vote share, the unit is an electoral constituency,

and the treatment is winning an election under plurality rules. We refer to this kind of RD design with multiple cutoffs as the Multi-Cutoff Regression Discontinuity Design.

When there are only two options or candidates in an election, the victory cutoff is always 50% of the vote, and it suffices to know the vote share of one candidate to determine the winner of the election and the margin by which the election was won. This occurs most naturally either in political systems dominated by exactly two parties or in elections such as ballot initiatives where the vote is restricted to only two yes/no options (e.g., DiNardo and Lee 2004). However, when there are three or more candidates, two races decided by the same margin might result in winners with very different vote shares. For example, in one district a party may barely win an election by 1 percentage point with 34% of the vote against two rivals that get 33% and 33%, while in another district a party may win by the same margin with 26% of the vote in a four-way race where the other parties obtain, respectively, 25%, 25%, and 24% of the vote.

The standard practice for dealing with this heterogeneity in the value of the cutoff has been to normalize the score so that the cutoff is zero for all units. For example, researchers often use the margin of victory for the party of interest as the running variable, defined as the vote share obtained by the party minus the vote share obtained by its strongest opponent. Using margin of victory as the score allows researchers to pool all observations together, regardless of the number of candidates in each particular district, and make inferences as in a standard RD design with a single cutoff. This normalizing-and-pooling approach is ubiquitous in political science and also in other disciplines. In section S2 of the appendix, we list several multi-cutoff RD examples in political science as well as in other fields, including education, economics, and criminology, where this approach has been applied.

Despite the widespread use of the normalizing-and-pooling strategy in RD applications, the exact form and interpretation of the treatment effect recovered by this approach has not been formally explored. Moreover, by normalizing and pooling the running variable, researchers may miss the opportunity to uncover key observable heterogeneity in RD designs, which can have useful policy implications. This is the motivation for our article. We generalize the conventional RD setup with a single fixed cutoff to an RD design where the cutoff is a random variable and use this framework to characterize the treatment effect parameter estimated by the normalizing-and-pooling approach. We show that the pooled parameter can be interpreted as a double average: the weighted average across cutoffs of the local average treatment effects for all units facing each particular cutoff value. This weighted average gives higher weights to those values of the cutoff that are most likely to occur and include more observations. Our derivations thus show that the pooled estimand is not equal to the overall average of the average treatment effects at every cutoff value, except under specific assumptions.

We also use our framework to characterize the heterogeneity that is aggregated in the pooled parameter, and the assumptions under which this heterogeneity can be used to learn about the causal effect of the treatment at different values of the score. Learning about RD treatment effects along the score dimension is useful for policy prescriptions. As we show, the probability of facing a particular value of the cutoff may vary with characteristics of the units. If these characteristics also affect the outcome of interest, then differences between treatment effects at different values of the cutoff variable may be due to inherent differences in the types of units that happen to concentrate around every cutoff value. However, if the cutoff value does not directly affect the outcomes and units are placed as if randomly at each cutoff value, then a treatment effect curve can be obtained.

We illustrate our results with three different RD examples based on vote shares, population, and test scores. The first example analyzes Brazilian mayoral elections in 1996–2012, following Klašnja and Titiunik (2016), and studies the effect of the Party of Brazilian Social Democracy (PSDB, Partido da Social Democracia Brasileira) winning an election on the probability that the party wins the mayor's office in the following election. The running variable is vote share, and the multiple cutoffs arise because there are many races with more than two effective parties. The second example is based on Brollo et al. (2013) and focuses on the effect of federal transfers on political corruption in Brazil, where transfers are assigned based on whether a municipality's population exceeds a series of cutoffs. The third example is based on Chay, McEwan, and Urquiola (2005), where school improvements are assigned based on past test scores, and the cutoffs differ by geographic region. Our examples illustrate the different situations that researchers may encounter in practice, including the important difference between cumulative and noncumulative multiple cutoffs, which we discuss in detail below.

After illustrating the main methodological results in the sharp multi-cutoff RD framework, we show how the main ideas and results for sharp RD designs extend to fuzzy RD designs, where treatment compliance is imperfect. Furthermore, in section S4 of the appendix, we discuss other extensions and results, covering a nonseparable RD model with unobserved unit-specific heterogeneity (Lee 2008), kink RD designs (Card et al. 2015), and the connections to multi-scores and geographic RD designs (Keele and Titiunik

2015; Papay, Willett, and Murnane 2011; Wong, Steiner, and Cook 2013). Finally, before concluding, we offer recommendations for practice to guide researchers in the interpretation and analysis of RD designs with multiple cutoffs.

MOTIVATION: RD DESIGNS BASED ON MULTIPARTY ELECTIONS

To motivate our multi-cutoff RD framework, we explore an RD design based on elections that studies whether a party improves its future electoral outcomes by gaining access to office (i.e., by becoming the incumbent party), a canonical example in political science. The treatment of interest is whether the party wins the election in year t, and the outcome of interest is the electoral victory or defeat of the party in the following election (for the same office), which we refer to as election at $t + 1$.

We apply this design to two different settings. First, we analyze US Senate elections between 1914 and 2010, pooling all election years and focusing on the effect of the Democratic Party's winning a Senate seat on the party's probability of victory in the following election for that same seat. Second, we analyze Brazilian mayoral elections for the PSDB between 1996 and 2012. We also pool all election years and focus on the effect of the party's winning office at t on the party's probability of victory in the following election at $t + 1$, which occurs four years later. For details on the data sources for the US and the Brazil examples see, respectively, Cattaneo, Frandsen, and Titiunik (2015) and Klašnja and Titiunik (2016).

Figure 1 presents RD plots of the effect of the party barely winning an election on the probability of victory in the following election in both settings, using the methods in Calonico, Cattaneo, and Titiunik (2015a). These figures plot the probability that the party wins election $t + 1$ (y-axis) against the party's margin of the victory in the previous election (x-axis), where the dots are binned means of binary victory variables, and the solid lines are fourth order polynomial fits. All observations to the right of the cutoff correspond to states/municipalities where the party won election t, and all observations to the left correspond to locations where the party lost election t. Figure 1A shows that, in Brazilian mayoral elections, the PSDB's bare victory at t does not translate into a higher probability of victory at $t + 1$. In contrast, as shown in figure 1B, a Democratic Party's victory in the Senate election at t considerably increases the party's probability of winning the following election at the cutoff for the same Senate seat.

For the statistical analysis of these two RD applications, we followed standard practice and used margin of victory as the score, thus normalizing the cutoff to zero for all elec-

tions. This score normalization is a practical strategy that allows researchers to analyze all elections simultaneously regardless of the number of parties contesting each electoral district or even across years. However, as we now illustrate, this approach pools together elections that are potentially heterogeneous.

If there were exactly two parties contesting the election in each state or municipality, the running variable or score that determines treatment would be the vote share obtained by the party at t, as this vote share alone would determine whether the party wins or loses election t. However, this is rarely the case in applications. For example, roughly 68% of US Senate elections and 50% of Brazilian mayoral elections are contested by three or more candidates in the periods for which we have data.[1]

While these two cases differ little in terms of the number of parties, the number of effective parties is quite different. In a race with three or more parties, in order to know whether a party's vote share led the party to win the election, and by how much, we need to know the vote share obtained by the party's strongest opponent—the runner-up when the party wins and the winner when the party loses. In the above example, if the Democratic candidate obtains 33.4% of the vote against two candidates that obtain 33.3% and 33.3%, its margin of victory is $33.4 - 33.3 = 0.1$ percentage points, and it barely wins the election. In contrast, when the other two parties obtain 60% and 6.6%, its margin of victory is $33.4 - 60 = -26.6$ points, and it loses the election by a large margin.

Figure 2 summarizes the strongest opponent's vote share for close elections in our two examples. Figure 2A shows the histogram of the vote share obtained by the PSDB's strongest opponent at election t only for races where the PSDB won or lost by three percentage points—that is, for races where the absolute value of the PSDB's margin of victory at t is 3 percentage points or less. Figure 2B shows the analogous figure for the Democratic Party in US Senate elections.

Figure 2 reveals that the degree of heterogeneity differs greatly between the two examples. In a perfect two-party system, the vote share of the party's strongest opponent in races decided by 3 percentage points or less would range from 51.5% to 48.5%. That is, 48.5% is the minimum vote percentage that a party could get in a two-party race if it lost to another party by a margin no larger than 3 percentage points; similarly, 51.5% is the maximum possible value. As

---

1. We use the terms "parties" and "candidates" interchangeably throughout, but we note that in US Senate elections some third candidates are unaffiliated with a political party.
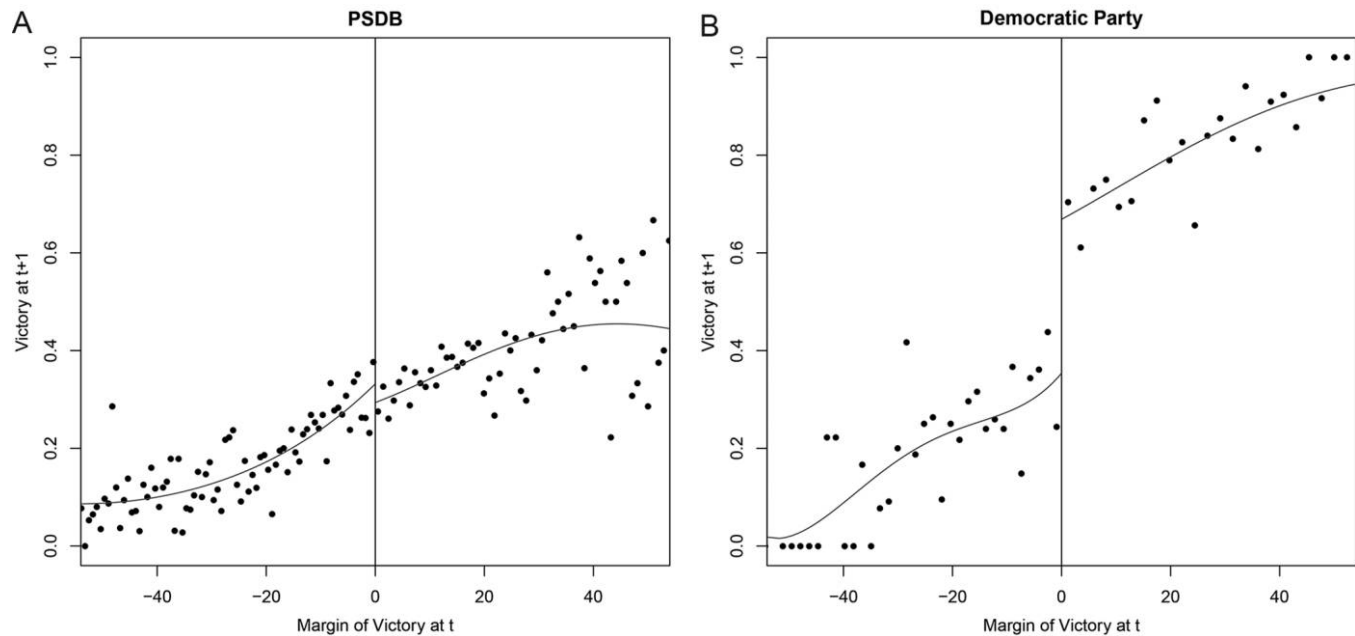
Figure 1. RD effect of party winning on party's future victory: Brazil and the United States. *A*, Brazilian mayoral elections, 1996–2012. *B*, US Senate elections, 1914–2010.

illustrated in figure 2*B*, in Senate elections where the Democratic Party wins or loses by less than 3 percentage points, only 23% of the observations are below 48.5%. Moreover, in 94% of the elections in the figure, the Democratic Party's strongest opponent gets 46% or more of the vote. Thus, despite most Senate elections having a third candidate, the

vote share obtained by such candidate is negligible in most cases, and there is little heterogeneity in the location of close races along the values of the strongest opponent's vote share.

In contrast, figure 2*A* shows that the PSDB exhibits much higher heterogeneity, with strongest opponent vote shares that fall below 48.5% for 46% of the observations. Moreover,
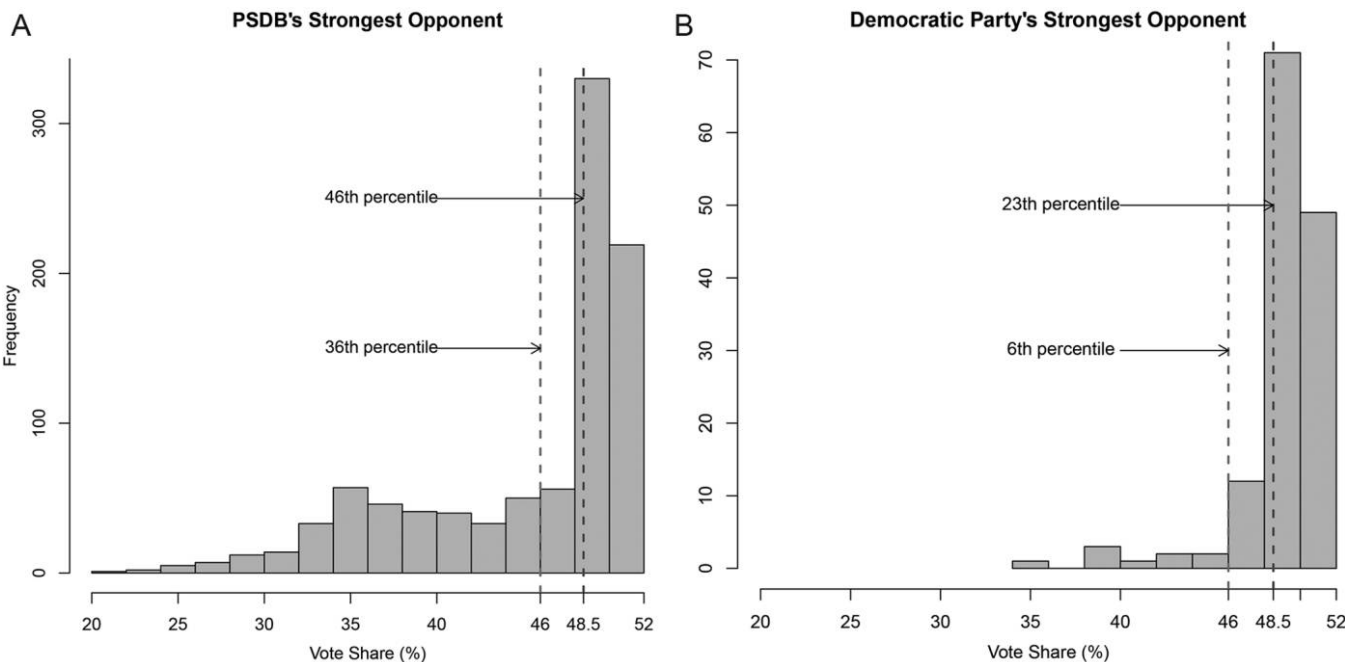

Figure 2. Histogram of vote share of strongest opponent in elections where the PSDB and the Democratic Party won or lost by less than 3 percentage points. *A*, Brazilian mayoral elections, 1996–2012. *B*, US Senate elections, 1914–2010.

more than a third of the elections (36%) have strongest opponent vote shares below 46%. In other words, a non-negligible proportion of the elections where the PSDB wins or loses by 3 points are elections in which third parties obtain a significant proportion of the vote. In particular, the histogram shows that in most races where the strongest opponent's vote share is less than 46%, smaller parties concentrate at least 17% to 20% of the vote.

The differences illustrated in figure 2 suggest that we ought to interpret the RD results in figure 1 differently. In the case of US Senate elections, the average effect in figure 1B can be interpreted as roughly the average effect of the Democratic Party barely reaching the 50% cutoff and thus winning a two-way race. Although the existence of third parties means that the real cutoff is not exactly 50%, in practice most close races are decided very close to this cutoff, so that the average RD effect can be roughly interpreted as the effect of winning at 50%. In contrast, the average effect in Brazil includes a significant proportion of elections where the cutoff is very far from 50%. As a consequence, this overall effect cannot be interpreted simply as the effect of barely winning at the 50% cutoff. Rather, it is the average effect of barely winning at different cutoffs that range roughly from 20% to 50% of the vote. For example, the PSDB may win an election by a 2-percentage-point margin, obtaining 51% of the vote against a single challenger who obtains 49%, or obtaining 36% of the vote against two challengers who get 34% and 30%. This heterogeneity is "hidden" or averaged in the normalizing-and-pooling strategy.

Importantly, the heterogeneity in the Brazil example is not unique or unusual. Many political systems around the world have third candidates who obtain a sizable proportion of the vote. Figure 3 shows the distribution of the vote share obtained by a reference party's strongest opponent in six different countries across different time periods and types of elections, using the data compiled by Eggers et al. (2015). These histograms show only the subset of races decided by less than 3 percentage points for legislative elections in Canada, the United Kingdom, Germany, India, New Zealand, and mayoral elections in Mexico—the reference party is indicated in each case. In all the elections illustrated in figure 3, there is a nonnegligible proportion of cases where the vote share of the party's strongest opponent falls below the range that would be observed in a perfect two-party system with 50% cutoff.

MULTI-CUTOFF REGRESSION DISCONTINUITY DESIGNS

We now we formally describe the heterogeneity in the treatment effect parameter that arises when the normalizing-and-

pooling approach is used in RD designs with multiple cutoffs. Our setup is general and applies to any running variable, not only vote shares. In this section, we discuss the interpretation of the pooled parameter, while the next section explores how to recover different quantities of potential interest under additional assumptions.

We study the sharp RD design first, and assume that the cutoff has finite support—that is, that it can only take a finite number of different values. We adopt these simplifications to ease the exposition, but we extend the framework to fuzzy multi-cutoff RD designs in the section Fuzzy Multi-Cutoff RD Designs below, and to kink RD designs and RD designs with multiple scores in section S4 of the appendix. Our assumptions and identification results reduce to those in Card et al. (2015), Hahn, Todd, and van der Klaauw (2001), an Lee (2008) for the special case of single-cutoff RD designs.

In the standard single-cutoff RD design framework, there are three key random variables for each unit, $(Y_i(0), Y_i(1), X_i)$ for $i = 1, 2, \ldots, n$, where $Y_i(0)$ and $Y_i(1)$ denote the potential outcomes for each unit when they are not exposed and exposed to treatment, respectively, and X_i denotes the running variable or score assigned to each unit. In a sharp RD setting, the treatment indicator for each unit is $D_i = \mathbb{I}(X_i \geq c)$, where c is a common known cutoff for all units and $1(\cdot)$ denotes the indicator function. The treatment effect of interest in this setting is the average treatment effect at the cutoff: $\tau = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = c]$. In this context, one can always assume $c = 0$, without loss of generality, by replacing X_i by $\widetilde{X}_i = X_i - c$ and taking $c = 0$ as the cutoff for all units.

In our multi-cutoff RD design framework, X_i continues to denote the running variable or score for unit i, but now there is another random variable, C_i, that denotes the cutoff that each unit i faces, which we assume has support $\mathcal{C} = \{c_1, c_2, \ldots, c_J\}$ with $\mathbb{P}[C_i = c] = p_c \in [0, 1]$ for $c \in \mathcal{C}$. We assume X_i is continuous with a continuous (Lebesgue) density $f_X(x)$, and let $f_{X|C}(x|c)$ denote a (regular) conditional density of $X_i|C_i = c$.[2] In the standard single-cutoff RD design, C_i would be a fixed value (i.e., $\mathbb{P}[C_i = c] = 1$), but in our framework it is a random variable taking possibly different values. As a result, it is possible for different units to face different cutoff values. In the motivating example based on Brazilian elections discussed above, the units indexed by i are municipalities, X_i is the vote share obtained by the PSDB, and C_i is the vote share of the PSDB's

---

2. Throughout the paper, we assume that all densities exist (with respect to the appropriate dominating measure) and are positive and that the Lebesgue densities are continuous at the evaluation points of interest.
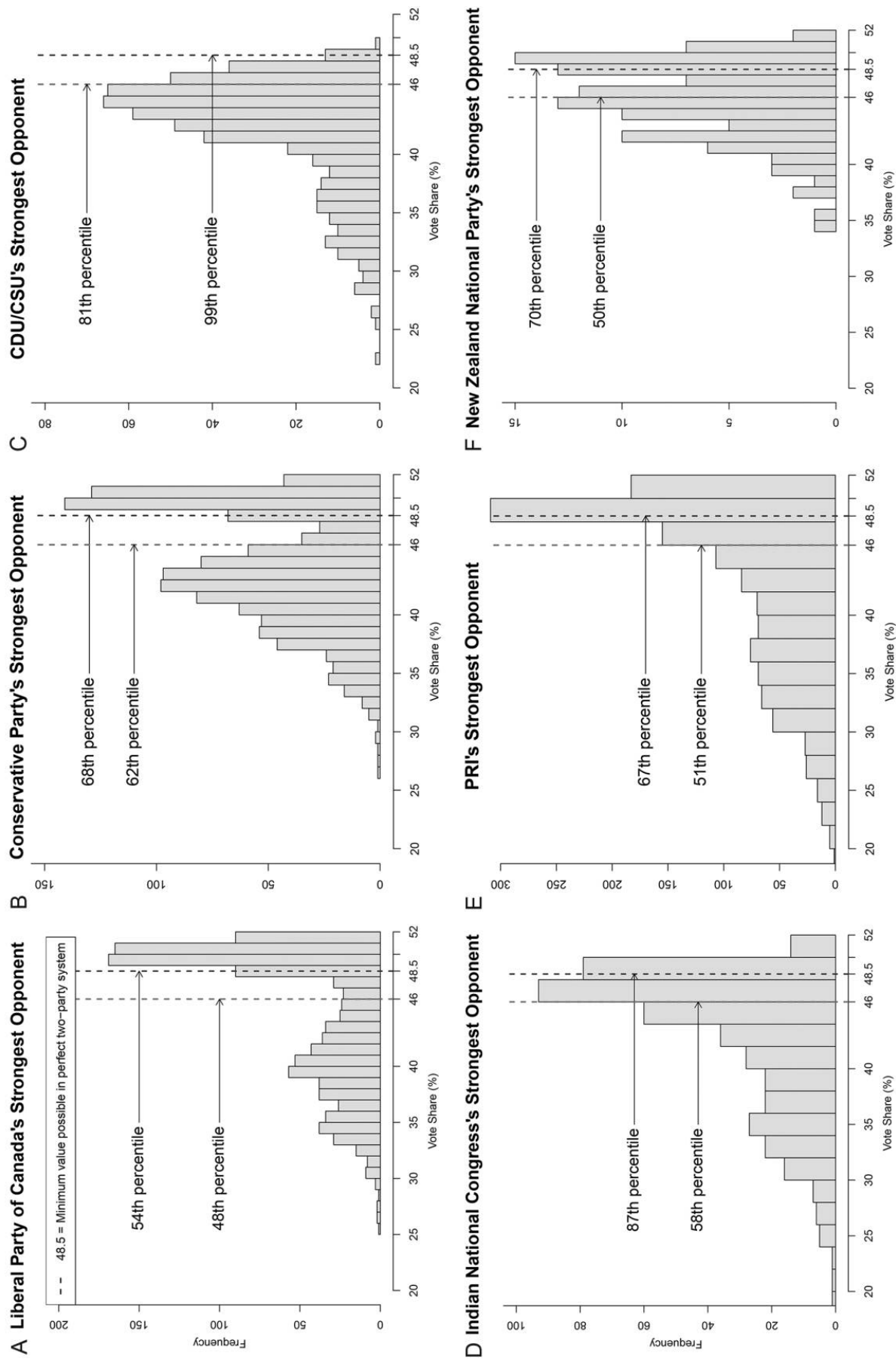
Figure 3. Strongest opponent's vote share in elections decided by less than 3 percentage points. *A*, Canadian House of Commons, 1867–2011. *B*, British House of Commons, 1918–2010. *C*, German Bundestag, 1953–2009. *D*, Indian Lower House, 1953–2009. *E*, Mexican municipalities, 1970–2009. *F*, New Zealand Parliament, 1946–1987. PRI = Institutional Revolutionary Party.

strongest opponent.[3] In the empirical illustrations we present below, X_i will be a vote share, a population measure, or a test score. The variable $D_i \in \{0, 1\}$ continues to be the treatment indicator, but now assignment to treatment depends on both the running variable and the cutoff C_i. The unit receives treatment if the value of X_i exceeds the value of the cutoff and receives the control condition otherwise, leading to $D_i = D_i(X_i, C_i) := \mathbb{I}(X_i \geq C_i)$. In the motivating examples discussed above, $D_i = 1$ when the party wins the t election in location i, and $D_i = 0$ if it loses. This setting captures perfect compliance or intention to treat; see Fuzzy Multi-Cutoff RD Designs for the more general fuzzy RD case.

A common practice in the context of RD designs with multiple cutoffs is to define the normalized running variable or score $\widetilde{X}_i := X_i - C_i$, pool all the observations as if there was only one cutoff at $c = 0$, and use standard RD techniques. In the motivating examples, $\widetilde{X}_i$ is the party's margin of victory at election t—that is, the party's vote share (X_i) minus the vote share of its strongest opponent (C_i)—and the party wins the election when this margin is above zero. That is, we can write $D_i = \mathbb{I}(\widetilde{X}_i \geq 0)$. It follows that the limit of D_i as X_i approaches $C_i = c$ from the left (i.e., from the region where $X_i \leq C_i$) is equal to zero, and it is equal to one when X_i approaches $C_i = c$ from the right. We formalize this in the assumption below, extended to the multi-cutoff RD setting.

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

$$\lim_{\varepsilon \to 0^+} \mathbb{E}[D_i | X_i = c + \varepsilon, C_i = c] = 1$$

and

$$\lim_{\varepsilon \to 0^+} \mathbb{E}[D_i | X_i = c - \varepsilon, C_i = c] = 0.$$

To complete the multi-cutoff RD model, we assume the observed outcome is $Y_i = Y_{1i}(C_i)D_i + Y_{0i}(C_i)(1 - D_i)$, where $Y_{1i}(c)$ and $Y_{0i}(c)$ are, respectively, the potential outcomes under treatment and control at each level $c \in \mathcal{C}$ for each unit $i = 1, 2, \ldots, n$. We employ the standard notation from the causal inference literature: $Y_{di}(C_i) = \sum_{c \in \mathcal{C}} \mathbb{I}(C_i = c)Y_{di}(c)$, for $d = 0,1$. Unlike the single-cutoff RD design, this model involves $2J$ potential outcomes, a pair for each cutoff level $c \in \mathcal{C}$. In our motivating examples, $Y_{1i}(c)$ is the party's victory or defeat that would be observed at election $t + 1$ if

the party won the previous election at t, and $Y_{0i}(c)$ is the party's victory or defeat that would be observed at the election if the party lost the previous election. Note that, for each state or municipality, we only observe $Y_{0i}(c)$ or $Y_{1i}(c)$, but not both, since the party cannot simultaneously lose and win election t. Instead, we observe Y_i a (binary) variable equal to one if the party wins election $t + 1$.

Our notation allows the cutoff for winning an election to affect the potential outcomes directly. More generally, the potential outcomes may be related to several variables: the running variable X_i, the cutoff C_i, and other unit-specific (unobserved) characteristics. The latter variables are usually referred to as the unit's "type"—see the supplemental appendix for further discussion. Thus, in our examples, we not only let the party's potential electoral success in election $t + 1$ be related to its vote share and the vote share of its strongest opponent at t, but also to other (potentially unobservable) characteristics of the state or municipality where the elections occur, such as its geographic location, the underlying partisan preferences of the electorate and its demographic makeup.

Finally, as is common in the RD literature, we assume that we observe a random sample, indexed by $i = 1, 2, \ldots, n$, from a well-defined population. As our notation also makes clear, we are explicitly ruling out interference between units; see, for example, Bowers, Fredrickson, and Panagopoulos (2013) and Sinclair, McConnell, and Green (2012), and references therein, for more discussion of SUTVA (Stable Unit Treatment Value Assumption) implications and violations in political science.

THE NORMALIZING-AND-POOLING APPROACH

The RD pooled estimand, τ^P, is defined as follows:

$$\tau^P = \lim_{\varepsilon \to 0^+} \mathbb{E}[Y_i | \widetilde{X}_i = \varepsilon] - \lim_{\varepsilon \to 0^+} \mathbb{E}[Y_i | \widetilde{X}_i = -\varepsilon]. \quad (1)$$

Equation (1) is the general form of the estimand in a multi-cutoff RD where the score has been normalized, all observations have been pooled, and the common cutoff is zero. Estimation of this pooled estimand is straightforward and, as discussed above, is done routinely by applied researchers. After normalization of the running variable, estimation just proceeds as in a standard RD design with a single cutoff—for example, using local nonparametric regression methods, as is now standard practice. We provide further details in the section Estimation and Inference in Multi-Cutoff RD Designs below. Although estimation of τ^P is straightforward, the interpretation of this estimand differs in important ways from the interpretation of the causal estimand in a standard single-cutoff RD design.

---

3. In multi-cutoff RD designs based on elections, C_i is a continuous random variable. As we illustrate in the section Empirical Examples, in order to analyze such examples within our framework, we discretize C_i by dividing its support into intervals.

We consider first the most general form of treatment effect heterogeneity where the treatment effect varies both across and within cutoffs. In this general case, individuals may respond to treatment differently if they face different cutoffs but also if they face the same one. Formally, this individual-level treatment effect is $\tau_i(c) = Y_{1i}(c) - Y_{0i}(c)$. In our motivational empirical example, this implies that the incumbency effect may vary in districts with different vote shares of the party's strongest opponent but it may also vary across districts with the same value of this variable. In order to derive the expression for τ^P we invoke the following two assumptions.

Assumption 2 (Continuity of Regression Functions). For all $c \in \mathcal{C}$: $\mathbb{E}[Y_{0i}(c)|X_i = x, C_i = c]$ and $\mathbb{E}[Y_{1i}(c)|X_i = x, C_i = c]$ are continuous in x at $x = c$.

Assumption 3 (Continuity of Density). For all $c \in \mathcal{C}$: $f_{X|C}(x|c)$ is positive and continuous in x at $x = c$.

Assumption 2 says that expected outcomes under treatment and control are continuous functions of the running variable at all possible cutoff values, implying that units barely below a cutoff are valid counterfactuals for units barely above it. This is the fundamental identifying assumption in all RD designs. Assumption 3 rules out discontinuous changes in the density of the running variable. Lemma 1 characterizes the pooled estimand under complete heterogeneity.

Lemma 1 (Pooled Sharp Multi-cutoff RD). If assumptions 1, 2, and 3 hold, the pooled sharp RD causal estimand is

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

$$\omega(c) = \frac{f_{X|C}(c|c)\mathbb{P}[C_i = c]}{\sum_{c \in \mathcal{C}} f_{X|C}(c|c)\mathbb{P}[C_i = c]}.$$

All proofs and related results are given in section S3 of the appendix. Lemma 1 says that whenever heterogeneity within and across cutoffs is allowed, the pooled RD estimand recovers a double average: the weighted average across cutoffs of the average treatment effects $\mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|X_i = c, C_i = c]$ across all units facing each particular cutoff value. Importantly, this derivation shows that the pooled estimand is not equal to the overall average of the (average) treatment effects at every cutoff value. In section S4.1 of the appendix, we discuss this point further and show the differences between the average of the cutoff-specific effects and τ^P and also discuss how the pooled estimand can be written as an average across individuals of different types as in Lee (2008).

Two things should be noted in order to interpret the estimand in lemma 1. First, the weight $\omega(c)$ determines the effects that are included in the pooled parameter τ^P and how much each effect contributes to this parameter. The term $\mathbb{P}[C_i = c]$ is simply the probability of observing the particular realization of each cutoff and implies that $\omega(c)$ will be higher for those values of c that are more likely to occur. The term $f_{X|C}(c|c)$ increases the weight of effects that occur at values of c where the density of the running variable is high.

Second, each of the conditional effects being averaged, $\mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|X_i = c, C_i = c]$, is the average effect of treatment given that both the running variable X and the cutoff C are equal to a particular value c. In the standard single-cutoff RD design, the effect recovered is the average effect of treatment at the point $X_i = c$, an effect that is typically characterized as local because it reflects the average effect of a treatment at a particular value of the running variable and is not necessarily generalizable to other values of X_i. Therefore, the conditional effects in the pooled RD case intensify the local nature of the effect, because they represent the average effect of treatment when both the running variable and the cutoff take the same particular value.

For example, in a perfect two-party system, the RD effect of a party winning election t on the party's future victory at $t + 1$ recovers a single effect—the effect of this party winning with a vote share just above 50%, not the effect of winning in general. In contrast, in the pooled RD design, this is just one of the effects that are included in τ^P. The pooled RD estimand τ^P includes other effects, such as the average of the party winning with 40% of the vote against a strongest opponent that gets just below 40%, the average effect of the party winning with 30% of the vote against a strongest opponent that gets just below 30%, and so forth. This heterogeneity in τ^P makes it a richer estimand, but it also makes each of its component effects more local or specific, because each reflects only one of the multiple ways in which "barely winning" can occur.

Moreover, τ^P is subtle in other ways. In the pooled multi-cutoff RD design, just like in the standard single-cutoff RD design, units whose score X_i is close to a cutoff may be systematically different from the units whose score is far from it. In the pooled RD design, however, units can also differ systematically in their probabilities of facing a particular value of the cutoff. For example, in the Brazilian mayoral context, municipalities where the PSDB gets 50% of the vote might be different in relevant ways from municipalities where the PSDB gets 35% of the vote. In addition, even

within those municipalities where the PSDB gets 35% of the vote, municipalities where the strongest opponent also gets roughly 35% may be very different from those where the strongest opponent gets 10% or 15% and the election is uncompetitive. In terms of our example, this means that, at every value c, the effects that contribute to τ^{P} are the average effect of the party barely defeating an opponent that obtained a vote share equal to c. While this effect is uninformative about the effects at other values of c, it does imply that when there are many values of c the pooled RD estimand contains information about the causal effect of barely winning in a number of different contexts. This aspect of the pooled RD estimand, by which many different local effects are combined when many different values of C_i may occur, shows that multi-cutoff RD designs contain a richer set of information relative to single-cutoff settings.

This means that the pooled estimand in a multi-cutoff RD design is something of a paradox. On the one hand, each of the cutoff-specific effects in τ^{P} is a very local parameter in the sense that it is the effect of the treatment for those units for which X_i barely exceeds C_i in only one of the multiple ways in which X_i could barely exceed C_i. On the other hand, when C_i takes a wide range of values, the average effect of treatment is recovered for the many different ways in which X_i can barely exceed C_i, potentially leading to a more global interpretation of the RD effect. We will use the two motivating examples, as well as two other distinct empirical illustrations, to illustrate how researchers may explore the richness in τ^{P}.

IDENTIFICATION IN MULTI-CUTOFF RD DESIGNS

A usual concern with single-cutoff RD designs is that they only offer estimates of the treatment effect at the cutoff and are thus uninformative about the magnitude of the treatment effect at other values of the running variable. In our motivating examples, the multi-cutoff RD gives us the effect of barely defeating the opponent party with a range of different values—in Brazil mayoral elections this range is roughly 20%–50%. Can we use this wider range of values to learn about a more global effect? We now consider assumptions under which the information contained in the pooled estimand τ^{P} can be disaggregated to learn about treatment effects of a more global nature.

Constant treatment effects

We first consider a simplification of the general case, where the treatment effect is different across cutoffs but constant for all individuals who face the same cutoff, that is, $\tau_i(c) := Y_{1i}(c) - Y_{0i}(c) = \tau(c)$ with $\tau(c)$ a fixed constant for all i facing the same c. Note that $\tau(c)$ varies by unit only insofar

as c varies by unit, but there is no i subindex in $\tau(c)$, indicating that two units facing the same given cutoff c will have the same treatment effect $\tau(c)$. In terms of our motivating examples, this assumption implies that the effect of the party winning an election on its future electoral success is the same in all municipalities/states where its strongest opponent obtains the same proportion of the vote. This is undoubtedly a very strong assumption. We include it here to illustrate one possible way in which the treatment effects recovered by the multi-cutoff RD design can be given a more global interpretation, but we discuss weaker assumptions in the subsequent sections.

The proposition below shows that when there is no heterogeneity within cutoffs, the relationship between the pooled RD estimand and the cutoff-specific effects simplifies considerably.

Proposition 1 (Constant Treatment Effects). Suppose the assumptions of lemma 1 hold. If $\tau_i(c) = \tau(c)$ for all i and $\tau(c)$ fixed for each c, then the pooled RD estimand is $\tau^{\mathrm{P}} = \sum_{c \in \mathcal{C}} \tau(c)\omega(c)$, where the weights $\omega(c)$ are the same as in lemma 1.

Thus, when effects are constant within cutoffs, $\tau(c)$ captures the effect of treatment for all individuals facing cutoff c. Naturally, proposition 1 simplifies considerably when the treatment effect is the same for all individuals at all cutoffs, that is, $\tau_i(c) = Y_{1i}(c) - Y_{0i}(c) = \tau$ for all i and all c, and thus $\tau(c) = \tau$ for all c. In this case, the pooled estimand becomes $\tau^{\mathrm{P}} = \sum_{c \in \mathcal{C}} \tau(c)\omega(c) = \tau \sum_{c \in \mathcal{C}} \omega(c) = \tau$, recovering the single (and therefore global) constant treatment effect. This global interpretation of the multi-cutoff RD estimand under constant treatment effects is analogous to the interpretation in a single-cutoff RD design, where the assumption of homogeneous treatment effects leads to the identification of the overall constant effect of treatment.

Ignorable running variable

The case introduced above is very restrictive, as it is natural to expect some heterogeneity in treatment effects among units facing the same value of the cutoff. We now consider the less restrictive case of unit-heterogeneity within cutoffs, but with an average treatment effect at every value of the cutoff that does not depend on the particular value taken by the score. We summarize this in the following assumption.

Assumption 4 (Score Ignorability). For all $c \in \mathcal{C}$:

$$\mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|X_i, C_i = c]$$
$$= \mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|C_i = c].$$

Under assumption 4, the running variable is ignorable once we condition on the value of the cutoff—that is, once the value of the cutoff is fixed, we assume that the average effect of treatment is the same regardless of the value taken by the score. The proposition below shows the form of the pooled RD estimand in this case.

Proposition 2 (Score-Ignorable Treatment Effects). Suppose the assumptions of lemma 1 hold. If assumption 4 holds, then the pooled RD estimand is

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

where the weights $\omega(c)$ are the same as in lemma 1.

Thus, when the average effect of treatment does not vary with the running variable X_i, $\mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|C_i = c]$ captures the effect of treatment for all values of X_i, not necessarily those that are close to the cutoff c. For example, $\mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|C_i = c]$ may reflect the average effect of the Democratic Party winning election t on its future electoral success for a given value of its strongest opponent's vote share, regardless of whether the party defeated its opponent barely or by a large margin. In this sense, the effects in proposition 2 are global in nature. Note, however, that the treatment effects are allowed to vary with the value of C_i, and therefore the expression for τ^P in proposition 2, although not necessarily local, is only averaging over the set of values that C_i can take, and the values of C_i that will be given positive weight are only those values where the density of X_i given C_i at $X_i = C_i = c$, $f_{X|C}(c|c)$, is positive. As such, τ^P still retains a local aspect.

Ignorable cutoffs

We now consider the case where the running variable is not ignorable but where the heterogeneity brought about by the multiple cutoffs can be restricted in ways that allow extrapolation. It is useful to introduce the analogy between the RD design with multiple cutoffs and an experiment that is performed in different sites or locations. In the latter case, internally valid treatment effect estimates from experiments in multiple sites are not necessarily informative about the effect that the treatment would have in a different site where the experiment has not been run. This means that the results from multi-site experiments may not allow researchers to extrapolate to the overall population, a concern that is not necessarily eliminated if the number of sites is large (Allcott 2015). The problem arises because the sites that are selected to run an experimental trial may differ from the overall population of sites in ways that are correlated with the

treatment effect. For example, sites where the treatment is expected to have large effects may be more likely to run experimental trials, leading to a positive "site selection bias" that would overestimate the effects that the treatment would have if it were implemented in the overall population. Alternatively, the population may differ across sites in a characteristic that is associated with treatment effectiveness (Hotz, Imbens, and Mortimer 2005). Of course, generalizing the treatment effect from one particular site to other locations can be done under additional assumptions.

Like in a multi-site experiment, in the multi-cutoff RD we have a series of internally valid estimates that we would like to interpret more generally. In the multi-site experiments literature, the strongest and simplest assumption under which the generalization of effects is possible is independence of locations with respect to potential outcomes. This condition is guaranteed by design when the units in the population are randomly assigned to different sites. In our context, we can make the analogous assumption that, conditional on the value of the running variable, the cutoff faced by a unit is unrelated to the potential outcomes. Formally, we can write this assumption as follows.

Assumption 5 (Cutoff Ignorability). For all $c \in \mathcal{C}$:

(a)
$$\mathbb{E}[Y_{1i}(c)|X_i, C_i = c] = \mathbb{E}[Y_{1i}(c)|X_i] \quad \text{and}$$
$$\mathbb{E}[Y_{0i}(c)|X_i, C_i = c] = \mathbb{E}[Y_{0i}(c)|X_i].$$
(b)
$$Y_{1i}(c) = Y_{1i} \quad \text{and} \quad Y_{0i}(c) = Y_{0i}.$$

Assumption 5*a* says that, conditional on the running variable, the potential outcomes are mean independent of the cutoff variable C_i. In addition, we need to ensure that the value of the cutoff does not affect the potential outcomes. This is equivalent to the "no macro-level variables" assumption in Hotz et al. (2005), and to the "local policy invariance" condition in Dong and Lewbel (2015). Assumption 5*b* formalizes the idea of an exclusion restriction, requiring that the cutoff level does not affect the potential outcomes directly.

To build intuition, note that if $c_0 \leq X_i < c_1$, then assumption 5 leads to $\mathbb{E}[Y_i|X_i, C_i = c_0] - \mathbb{E}[Y_i|X_i, C_i = c_1] = \mathbb{E}[Y_{1i} - Y_{0i}|X_i]$ for observed random variables, which captures the average treatment effect conditional on X_i for $c_0 \leq X_i < c_1$. This shows that under these assumptions we can estimate the average treatment effect away from the cutoff and thus obtain a more global effect. However, the following lemma shows that, as before, the ability to recover

a global effect from the pooled multi-cutoff RD design, even under assumption 5, is limited by the fact that τ^P weighs these average effects by the probability of observing a realization of the cutoff variable C_i at the particular value c.

Proposition 3 (Cutoff-Ignorable Treatment Effects).
Suppose the assumptions of lemma 1 hold. If assumptions 5 holds, then the pooled RD estimand becomes

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

where the weights $\omega(c)$ are the same as in lemma 1.

Thus, under these assumptions, τ^P averages the average treatment effects $\mathbb{E}[Y_{1i} - Y_{0i}|X_i = c]$, each of which is the average effect of receiving treatment conditional on the running variable X_i being at the value c, regardless of the value taken by C_i. In our motivating examples, this represents the average effect of a party winning the t election given that the party's vote share is c and regardless of the vote share obtained by its strongest opponent, that is, regardless of whether it won barely or by a large margin. However, these averages are still evaluated only at values of c that are in the support of the random cutoff variable C_i. So, although they are more global effects, they can only be recovered at feasible values of C_i. Moreover, the weights $\omega(c)$ entering τ^P still depend on $\mathbb{P}[C_i = c]$.

If, in addition to the assumptions imposed in proposition 3, we impose the assumption that the conditional density of the score X_i given C_i is constant, the pooled RD parameter τ^P simplifies to:

$$\tau^p = \sum_{c \in \mathcal{C}} \mathbb{E}[Y_{1i} - Y_{0i}|X_i = c]\mathbb{P}[C_i = c]$$

and now, if the support of C_i is equal to the support of X_i (which will only be possible if both are discrete or both are continuous), we can recover the average of the average treatment effect at all values of X_i determined by the cutoff values faced by the units in the sample. All these assumptions combined would thus make τ^P a truly global averaged estimand, without the need of imposing an assumption of constant RD treatment effects.

Assumption 5 also has another important application. Under the conditions imposed in that assumption, $\mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|X_i = c, C_i = c] = \mathbb{E}[Y_{1i} - Y_{0i}|X_i = c]$. This shows that when these assumptions hold, estimating the RD effects separately for each value c will provide a treatment-effect curve that will summarize the effects of the treatment at different values of the running variable (independently of the value taken by the cutoff). In other words, under these assumptions,

we can estimate multiple RD treatment effects for different values of the running variable.

Of course, assumption 5 is generally strong and may be too restrictive in some empirical applications. In research under way, we are investigating different approaches in a multi-cutoff RD design to achieve identification of $\mathbb{E}[Y_{1i} - Y_{0i}|X_i = x, C_i = c]$ for values $x \neq c$ under substantially weaker conditions than assumption 5. These alternative conditions would allow for "endogenous cutoffs" or "sorting into cutoffs" for the units of analysis and would give an opportunity for extrapolation of RD treatment effects in applications where there is variation in cutoff values.

Difference between noncumulative and cumulative cutoffs

The plausibility of the assumptions just discussed will be directly affected by the way in which the multiple cutoffs are related to the running variable. Multi-cutoff RD designs are typically of two main types. In the first type, the value of the running variable X_i and the cutoff variable C_i are unrelated, in the sense that a unit i with running variable equal to a particular value, say $X_i = x_0$, can be exposed to any cutoff value $c \in \mathcal{C} = \{c_1, c_2, \ldots, c_J\}$. This scenario, which we call the Multi-cutoff RD Design with noncumulative cutoffs, is illustrated in figure 4A. As shown in panels I, II, and III, a unit with $X_i = x_0$ can be exposed to any one of the possible cutoff values—c_0, c_1, or c_2. In this scenario, the rule that governs whether a unit faces c_0, c_1, or c_2 may be related to X_i, but this rule is not a deterministic function of X_i.

RD designs based on multiparty elections have noncumulative cutoffs. For example, when the PSDB contests an election against two other parties, if it obtains 40% of the vote, its strongest opponent's vote share—the cutoff the PSDB faces to win the election—can be anything between 60% (if the third party gets zero votes) and just above 30% (if the second and third parties are tied). Thus, except for the restriction that the total sum of vote percentages must be 100%, the cutoff faced by the PSDB is unrelated to the vote share it obtains.

In contrast, some multi-cutoff RD applications have what we call cumulative cutoffs. In these applications, different versions of the treatment are given for different ranges of the running variable, and as a result the cutoff faced by a unit is a deterministic function of the unit's score value. In the hypothetical example illustrated in figure 4B, units with $X_i < c_0$ receive treatment A, units with $c_0 \leq X_i < c_1$ receive treatment B, units with $c_1 \leq X_i < c_2$ receive treatment C, and units with $c_2 \leq X_i$ receive treatment D. Thus, knowing a unit's score value is sufficient to know which cutoff (or pair of cutoffs)
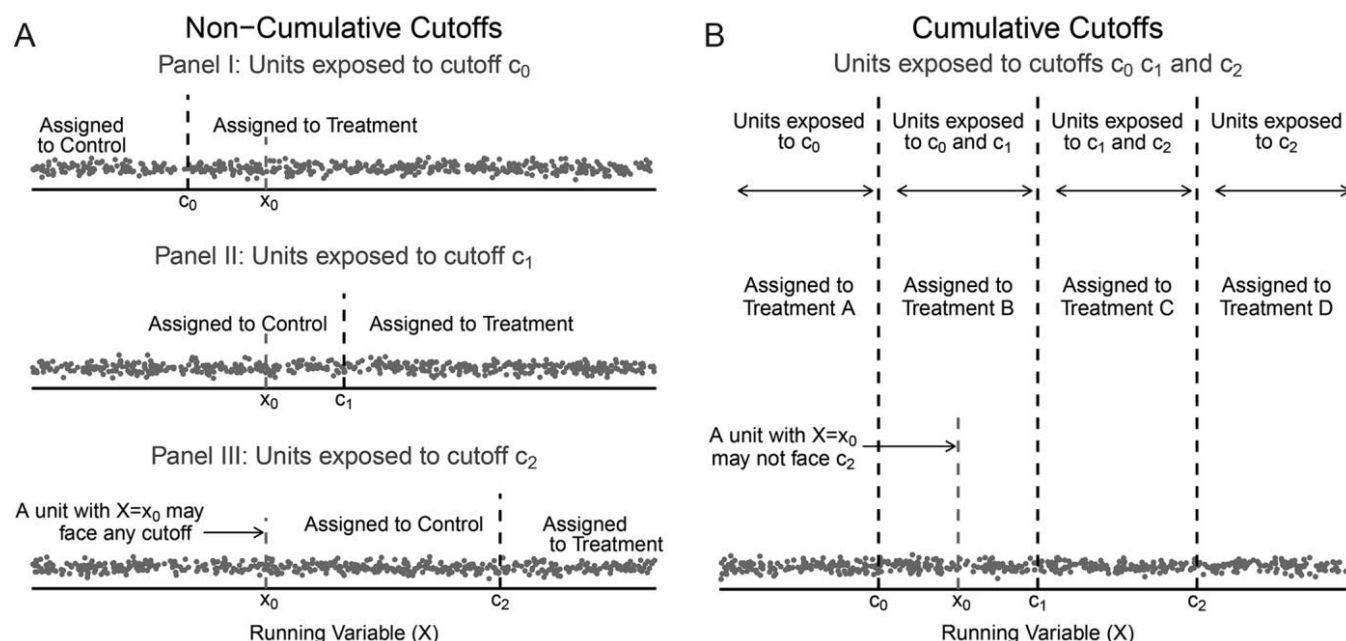
Figure 4. Cumulative versus noncumulative cutoffs in multi-cutoff RD designs: under noncumulative cutoffs, all units may be exposed to all cutoffs regardless of their score value; under cumulative cutoffs, units with a given score may be exposed to only a subset of cutoffs.

the unit faces. For example, an education intervention that gave a financial award to teachers based on evaluation scores could grant no awards to teachers with score below c_0, a small award to teachers with scores between c_0 and c_1, a medium award to teachers with scores between and c_2, and the largest awards to those whose evaluation scores are above c_2.

The difference between noncumulative and cumulative cutoffs is important for two main reasons. First, in designs with noncumulative cutoffs, all units tend to receive the same treatment, while in designs with cumulative cutoffs the treatments given are typically different in some respect. For example, a party whose vote share barely exceeds its strongest opponent's vote share always wins the election regardless of how low or high the strongest opponent's vote share is, while a teacher's award can be smaller or larger depending on which cutoff the teacher's score exceeds. This distinction may not be important if, in cumulative cutoff applications, researchers are willing to redefine the treatment appropriately. For example, all teachers see an increase in the award amount when they barely exceed any cutoff, and thus the treatment can be understood as increasing the award amount, regardless of by how much.

Second, while all of our results apply to both scenarios, the interpretation and plausibility of the underlying assumptions will change depending on whether a cumulative or noncumulative setting is considered. For example, our main lemma 1 applies to both cases, which means that regardless of whether cutoffs are cumulative or noncumulative, the normalizing-and-pooling approach leads to a weighted

average of cutoff-specific effects. However, the assumption of Cutoffs Ignorability (assumption 5) is less plausible under cumulative cutoffs because the cumulative rule implies a complete lack of common support in the value of the running variable for units facing different cutoffs. For example, in figure 4B, a unit with score $X_i = x_0$ can only be exposed to cutoff c_1 or c_0 but will never be exposed to c_2, and the units exposed to c_2 have score $X_i \geq c_1$, meaning that there are no units with low values of X_i exposed to c_2. In general, with cumulative cutoffs, the subpopulations of units exposed to every cutoff will have systematically different values of the running variable. Thus, if the running variable is related to the potential outcomes, the assumption that the potential outcomes are mean independent of the cutoff variable conditional on the running variable will always be false. The conditions required to obtain more general estimands based on multiple cutoffs are therefore much stricter in cases where the cutoffs are cumulative. We return to this distinction when we discuss our three empirical examples below.

ESTIMATION AND INFERENCE IN MULTI-CUTOFF RD DESIGNS

Estimation and inference in multi-cutoff RD designs can be based on the same methods and techniques that are commonly used for the analysis of single-cutoff RD designs, by either pooling all observations via a normalized score (as commonly done in current practice) or by conducting inference procedures for each cutoff separately. For a review of the most recent single-cutoff RD approaches to estima-

tion and inference, see Skovron and Titiunik (2016) and references therein.

The standard practice in single-cutoff RD analysis is to employ either local polynomial methods (Calonico, Cattaneo, and Farrell 2016; Calonico, Cattaneo, and Titiunik 2014a, 2014b, 2015b; Hahn et al. 2001) or local randomization methods (Cattaneo et al. 2015; Cattaneo, Titiunik, and Vazquez-Bare 2016a, 2016b; Lee 2008). Either approach can be used directly in multi-cutoff RD designs, both when a single normalized cutoff is considered ($\widetilde{X}_i = X_i - C_i$ and cutoff $c = 0$) or when the different cutoffs are analyzed separately (X_i and cutoffs $c \in \mathcal{C}$). We illustrate both approaches below with our three empirical illustrations, which cover both noncumulative and cumulative cutoff settings.

We briefly outline the main steps for estimation and inference using nonparametric local polynomial methods, which are usually the preferred option in empirical work. In this setting, point estimation amounts to fitting a weighted least squares regression of the outcome (Y_i) on a polynomial basis of the running variable ($\widetilde{X}_i$ or X_i) for observations within a small region around the cutoff ($c = 0$ when $\widetilde{X}_i$ is used or for each $c \in \mathcal{C}$ when X_i is used). The region around the desired cutoff is determined by a choice of bandwidth, and it is necessarily different depending on whether normalizing and pooling is used or not. The weights are determined by a kernel function, and the polynomial is fitted separately for observations above and below the cutoff. The RD treatment effect is obtained as the difference in the intercepts of the two polynomial fits at the cutoff(s), which implies that either one single estimate is computed ($\hat{\tau}^{\mathrm{P}}$ when normalizing and pooling) or a collection of estimates is computed ($\hat{\tau}^{\mathrm{P}}(c)$ for $c \in \mathcal{C}$ when X_i is used). The implementation of this procedure requires a bandwidth, which is typically chosen to minimize an approximation to the asymptotic mean-squared-error (MSE) of the point estimator(s). Confidence intervals for each parameters $\hat{\tau}^{\mathrm{P}}$ or (c), $c \in \mathcal{C}$, can be constructed using the asymptotically valid procedures developed by Calonico et al. 2014b), which have better finite-sample properties and faster vanishing coverage error rates.

Thus, implementing local polynomial estimation and inference in multi-cutoff RD designs is straightforward. By construction, the normalizing and pooling treats the multi-cutoff RD design as a single-cutoff RD design for all practical purposes, and all results in the literature are directly applicable. Likewise, a cutoff-by-cutoff analysis of multi-cutoff RD designs can also be done with estimation and inference methods already available in the literature with minor modifications and extra care. If the cutoffs are noncumulative and the cutoff is a discrete random variable, for every cutoff $c \in \mathcal{C} = \{c_1, c_2, \ldots, c_J\}$, researchers can con-

struct point estimators, confidence intervals, and other inference procedures by first keeping only the observations exposed to cutoff c and then employing directly local polynomial methods treating the cutoff c as the single cutoff in this subsample. When the cutoffs are noncumulative but continuous, as in the case of multiparty elections, we have $C_i \in [c_{\min}, c_{\max}]$. In this case, the researcher can first define a grid of J values $\mathcal{C} = \{c_1, c_2, \ldots, c_J\}$, in the interior of the support of the continuous cutoff, $[c_{\min}, c_{\max}]$, keep observations in a region around each grid value c_j, $j = 1, 2, \ldots, J$, and perform estimation and inference in each subsample treating each grid point c_j as the single cutoff. We illustrate this approach in our first empirical illustration below.

A similar procedure can be applied when the cutoffs are cumulative, either discrete or continuous, except that in this case the observations used for estimation and inference at each cutoff or grid point $c_j \in \mathcal{C}$ should only include observations whose running variable is not smaller than the cutoff immediately before and no larger than the cutoff immediately after c_j. For example, a reasonable empirical practice to analyze cutoff c_j would be to consider only observations with score variable satisfying $c_{j-1} + k_{j-1} < X_i < c_{j+1} - k_{j+1}$, assuming the cutoffs are ordered, where k_{j-1} and k_{j+1} could be chosen to the middle point or the median point (based on X_i) between c_{j-1} and c_j, and between c_j and c_{j+1}, respectively.

In all cases, the individual point estimates and confidence intervals can then be plotted against each cutoff or grid value in $\mathcal{C} = \{c_1, c_2, \ldots, c_J\}$ to capture the heterogeneity underlying the pooled RD treatment effect τ^{P}. Joint inference across different cutoffs is also possible by either relying on the bootstrap or by deriving the joint asymptotic distribution of the cutoff-specific estimates.

EMPIRICAL EXAMPLES

We now illustrate how the formal results derived above can inform the empirical analysis of RD designs with multiple cutoffs. We analyze three different examples: the incumbency advantage example in Brazil presented above, the effect of federal transfers on political corruption in Brazil analyzed by Brollo et al. (2013), and the effect of school infrastructure improvements on educational outcomes analyzed by Chay et al. (2005). We do not analyze the US Senate example further because the number of the effective parties is very close to two—see section S5 in the appendix for more details.

Example 1: The effect of incumbency for the PSDB in Brazilian elections

The first example we analyze is the PSDB's incumbency advantage in Brazilian mayoral elections introduced above.

In this electoral context, about a third of races occurs in municipalities where the two top-getters combined obtain less than 70% of the vote. Table 1 presents the frequency of races in our sample by different levels of the PSDB's strongest opponent vote shares at t. Since this variable is continuous, we divide its support in four nonoverlapping intervals: [0, 35), [35, 40), [40, 45), and [45, 50). Within each of these intervals, table 1 reports the number of elections that the party won and lost at t. 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 multiparty RD design. For example, the PSDB wins less than 64% of the races where the vote share of its strongest opponent is 35% or higher.

We begin by estimating τ^P, which is the pooled RD estimand that uses margin of victory as the score and normalizes all cutoffs to zero, by local linear regression with MSE-optimal bandwidth. The pooled RD point estimate is -0.036, an effect that cannot be statistically distinguished from zero at conventional levels (robust p-value $= .144$). The robust 95% confidence interval is $[-0.110, 0.016]$.

Next, we explore the heterogeneity by separately estimating the RD effects at different levels of strongest opponent's vote share. We choose a grid of values in the support of the vote share of the PSDB's strongest opponent and, for each value in this grid, we separately estimate the RD effect of the PSDB's winning at t on the PSDB's future success using only the 600 treated observations closest to the grid value and the 600 control observations closest to the grid value.

Figure 5 summarizes the results, showing the treatment effects at six different, equidistant values of strongest opponent vote shares between 34% and 49%. The dots are the treatment effects and the bars are the robust 95% confidence intervals described in the section Estimation and

Table 1. Frequency of Observations for Different Levels of the PSDB's Strongest Opponent Vote Shares at t

PSDB in Brazil Mayoral Elections			
Strongest Opponent Vote (%) (Cutoff Value)	Sample Size	Victories (%)	Defeats (%)
[0, 35)	1,346	84.9	15.1
[35, 40)	986	63.9	36.1
[40, 45)	1,251	62.3	37.7
[45, 50)	1,490	61.5	38.5

Note. Counts based on mayoral elections in Brazil in 1996–2012. Source is Klašnja and Titiunik (2016) replication data.
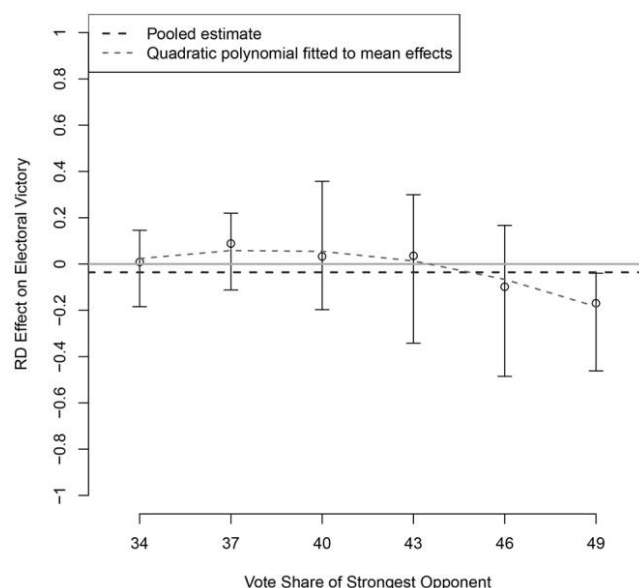

Figure 5. RD effects of PSDB's victory on future vote share at different levels of strongest opponent's vote share.

Inference in Multi-Cutoff RD Designs. Note that for every value of the PSDB's strongest opponent vote share that is displayed in the figure, we are estimating the effect of the PSDB's barely defeating its strongest opponent, so that all the effects in this figure are local RD effects. The blue dotted line indicates the normalizing-and-pooling point estimate, $\hat{\tau}^P = -0.036$.

The effects shown in this figure reveal some heterogeneity. For values of strongest opponent vote shares that fall near 46% or below, the effect of barely winning is relatively small and cannot be distinguished from zero. This estimate is also consistent with the results from the pooled analysis. However, for those elections where the PSDB's strongest opponent obtains a vote share near 49%, the effect is negative and significantly different from zero at the 5% level.

The heterogeneity illustrated in figure 5 must be interpreted with care for two reasons. The first reason is practical. As shown in table 1, the number of observations at every cutoff is moderate, which may lead to noisy estimates of the conditional expectations. The length of the confidence intervals in figure 5 varies significantly across the range of the running variable, often increasing where the density of observations is lower.

Second, following our discussion in the section Estimation and Inference in Multi-Cutoff RD Designs, the interpretation of the treatment-effect curve in figure 5 depends crucially on the assumptions surrounding the factors that affect the strongest opponent's vote shares. If we were willing to assume that, at every level of vote share obtained by the PSDB at t, the vote share obtained by its strongest opponent

is mean independent of the PSDB's potential victory at $t + 1$ (assumption 5a) and the strongest opponent's vote shares affect the potential future performance of the PSDB only through the PSDB's winning or losing the election but not directly (assumption 5b), then each of these effects would be the effect of the PSDB winning election t with a vote share in each interval, regardless of whether it won barely or by a large margin.

If however, we believe that the more plausible scenario is one in which elections that differ in the strongest opponent's vote share also differ systematically in observed and unobserved factors that affect the PSDB future vote shares (e.g., municipalities with strong third parties may be systematically different from municipalities where only two parties contest the election), then the interpretation of figure 5 changes considerably. Under this scenario, the potential differences between the effects also reflect the different electoral environments that occur at different levels of strongest opponent's vote shares, and cannot be simply interpreted as the effect of treatment at those levels of the PSDB's t vote share (the running variable).

Example 2: The effect of federal transfers on political corruption in Brazil

Our second empirical illustration is based on a study by Brollo et al. (2013), who examine whether increasing federal transfers results in increased political corruption in Brazilian municipalities. Brazilian municipal governments provide goods and services related to education, health, and infrastructure. For municipalities with a population of less than 50,000 people, the largest source of total revenues is the Fundo de Participação dos Municípios (FPM) which are automatic transfers from the central government. FPM transfers are based on the population of the municipality within each state, increasing at preset population thresholds.

In the original study, the authors focused on the first seven thresholds: 10,189, 13,585, 16,981, 23,773, 30,565, 37,357, and 44,149. At each of these thresholds, the amount of FPM transfers increased by a linear multiplier. The question of interest is whether these increases in revenues contributed to political corruption, measured in various ways. In our reanalysis, we focus on a single corruption measure—a binary outcome equal to one if authorities found evidence of severe irregularities in municipal finances, including diversion of funds, over-invoicing of goods and services, and fraud.

The original study treated the design as a fuzzy RD, since the theoretical transfers that a municipality should receive based on official population counts are not always equal to the actual amount of FPM transfers received. This non-compliance arises from several sources, including the fact

that FPM transfer amounts were frozen for several years, while population counts shifted. We only focus on the intention-to-treat effects of population on corruption and thus analyze the data as a sharp RD design where population is the running variable and the treatment is having a population count that exceeds the cutoff for an increase in FPM transfers.

This application is an example of a multi-cutoff RD design with cumulative cutoffs: municipalities of a certain population are only exposed to one or at most two cutoffs, and the treatment assigned differs at different cutoffs, as being above each of the cutoffs results in a different amount of FPM transfers. For example, a municipality above the 30,565 cutoff receives more federal transfers than a municipality above the 16,981 cutoff. The treatment received at every cutoff is therefore changing, which is typical of cumulative cutoff settings.

The pooled RD point estimate in this application is 0.149 (robust p-value = .073), and the robust 95% confidence interval is $[-0.017, 0.375]$. We also estimate cutoff-by-cutoff effects. Figure 6, which is analogous to figure 5, shows the RD treatment effects at each of the seven different cutoffs. For every cutoff-specific effect, we only use observations with score greater than or equal to the previous cutoff, and smaller than the following cutoff. That is, at each cutoff c_j, we only include in the estimation observations with $c_{j-1} \leq X_i < c_{j+1}$, and at the extreme cutoffs, c_1 and c_J, we keep, respectively, observations with $X_i < c_2$ and $X_i \leq c_{J-1}$. The sample size at each cutoff is shown in table 2.

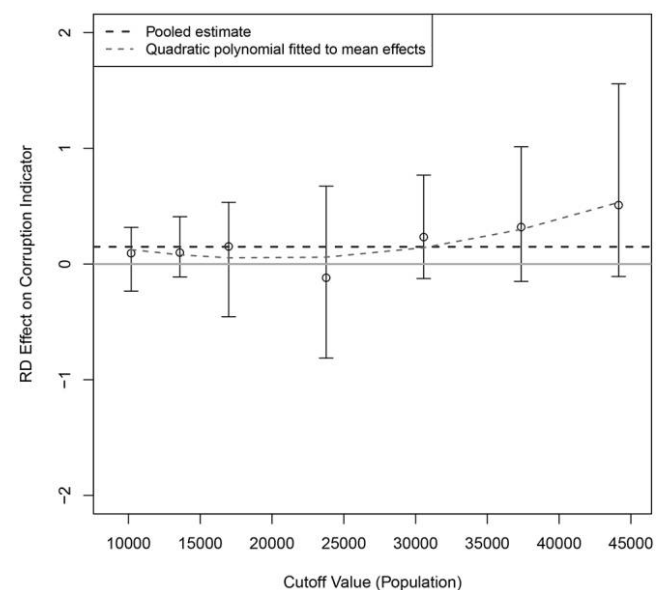As shown in figure 6, most point estimates are positive and near the pooled effect, although the effect at the last

Figure 6. RD effects of municipal transfers on corruption

Table 2. Frequency of Observations Exposed to Each Cutoff Value

Brazilian Municipalities	
Population (Cutoff Value)	Sample Size
10,189	489
13,585	432
16,981	407
23,773	342
30,565	225
37,356	153
44,148	81

Note. For each cutoff, the sample size is municipalities with score greater than or equal to previous cutoff (if there is one) and smaller than the following cutoff (if there is one). Source is Brollo et al. (2013) replication data.

population cutoff is considerably larger (but also highly variable due to the small number observations). The robust 95% and 90% confidence intervals for each cutoff-specific effect include zero. Since the 90% confidence intervals for the pooled effect do exclude zero, this suggests that the normalizing-and-pooling approach leverages the increased statistical power obtained by aggregating the sample sizes across all cutoffs.

Example 3: The effect of school infrastructure improvements in Chile

Our third and final empirical illustration is based on the study by Chay et al. (2005) on the effect of school improvements on test scores, with cutoffs that differ by geographic region. In 1990, the Chilean government introduced P-900, an intervention targeted at low-performing, publicly funded schools. Schools selected for participation in the P-900 intervention received improvements in their infrastructure, updated instructional materials, additional teacher training, and new after-school tutoring sessions. Assignment to P-900 participation was done using a single score based on a combination of school-level test scores in language and mathematics in 1988.[4] However, officials from the Chilean Ministry of Education used different cutoffs across each of Chile's 13 administrative regions, the highest subnational level of government.

_____

4. While the indicator for participation in P-900 and the test scores that make up the running variable are fully observed, the exact cutoffs in the score are not observed. Chay et al. (2005) use two different methods to estimate the cutoffs. We use the second set of estimated cutoffs in our analysis.

Table 3 contains the cutoff, number of observations, and range of the running variable in each region for the sample of urban, larger schools originally analyzed by the authors.[5] The outcome variables are school-level test score gains between 1988 and 1992 in language and mathematics. To keep our analysis brief, we focus only on language test score gains.

As in the other empirical examples, we first estimate the single pooled estimate of the effect of the P-900 intervention on language test scores. This estimate is 2.83 (robust p-value $= .003$), with 95% robust confidence interval [1.14, 5.44]. Thus, the normalizing-and-pooling strategy indicates that the program increased language test scores by nearly 3 points, an effect that is significantly different from zero at conventional levels.

We also explore whether the effect of the program varied by region. As in our previous application, the size of the subpopulations exposed to each cutoff value is very variable. For example, regions 11 and 12, which have the same threshold, include only 44 schools combined, while three other regions have roughly 500 or more schools. Because of the small number of observations, we exclude regions 11 and 12. The effects at all other cutoffs are presented in figure 7. The RD effects at four of the cutoff values are positive, and two of those are significantly different from zero at the 5% level. Two effects have negative point estimates, but the small number of observations prevents us from distinguishing these effects from zero and leads to large confidence intervals, particularly at the smallest cutoff. All in all, the effects of the P-900 program on language score gains seem to be moderately heterogeneous, although we must interpret this heterogeneity cautiously due to the variability of the cutoff-specific effects.

FUZZY MULTI-CUTOFF RD DESIGNS

All the results presented above can be extended in multiple ways. In this section we briefly discuss the fuzzy multi-cutoff RD design, where treatment compliance is imperfect. In section S4 of the appendix we further extend our work to the case of kink multi-cutoff RD designs and discuss connections with RD designs with multiple running variables.

In the fuzzy RD case, some units below the cutoff may receive the treatment and some units above it may refuse it, leading to a jump in the probability of receiving treatment at the cutoff that is less than one. Despite the necessary technical modifications, all the conceptual issues discussed above apply directly to this case. Therefore, for brevity, we only discuss here the interpretation of the pooled estimand.

_____

5. The schools included are urban schools with 15 or more students in the fourth grade in 1988.

Table 3. Cutoffs and Sample Sizes with Regions with Same Cutoff Values Combined

Geographic Region	Past Test Scores Index (Cutoff Value)	Sample Size	Min X_i	Max X_i
Region 7	42.4	157	33.62	81.55
Regions 6, 8	43.4	497	28.87	80.87
Region 13	46.4	959	31.00	83.35
Region 9	47.4	197	31.96	81.25
Regions 2, 5, 10	49.4	560	35.49	83.53
Regions 1, 3, 4	51.4	190	33.55	82.23
Regions 11, 12	52.4	44	40.75	82.65

Note. For each cutoff, the sample size is the number of schools in each region facing a unique cutoff. Source is Chay, McEwan, and Urquiola (2005) replication data.

First, we formalize the idea of imperfect treatment compliance in the multi-cutoff RD design.

Assumption 6 (Fuzzy RD). For all $c \in \mathcal{C}$:

$$\lim_{\varepsilon \to 0^+} \mathbb{E}[D_i | X_i = c + \varepsilon, C_i = c]$$
$$\neq \lim_{\varepsilon \to 0^+} \mathbb{E}[D_i | X_i = c - \varepsilon, C_i = c].$$

This assumption is a direct generalization of assumption 1 and covers as a special case the sharp RD design. Observe that D_i continues to denote whether unit i received treatment or not, but it is no longer required that this binary indicator take the form $D_i = \mathbb{I}(X_i \geq C_i)$ as in the sharp RD case.

The pooled estimand in the fuzzy RD design is generalized to

$$\tau_{\text{FRD}}^{\text{P}} = \frac{\lim_{\varepsilon \to 0^+} \mathbb{E}[Y_i | \widetilde{X}_i = \varepsilon] - \lim_{\varepsilon \to 0^+} \mathbb{E}[Y_i | \widetilde{X}_i = -\varepsilon]}{\lim_{\varepsilon \to 0^+} \mathbb{E}[D_i | \widetilde{X}_i = \varepsilon] - \lim_{\varepsilon \to 0^+} \mathbb{E}[D_i | \widetilde{X}_i = -\varepsilon]}.$$

The extension to fuzzy designs can be studied using a causal inference framework (Angrist, Imbens, and Rubin 1996), with a few simple modifications. Let the function $D_{0i}(x; c) : (-\infty, c) \times \mathcal{C} \to \{0, 1\}$ denote the potential treatment status when unit i faces cutoff c and has a score of $x < c$. Similarly, define the function $D_{1i}(x; c) : [c, \infty) \times \mathcal{C} \to \{0, 1\}$ as the potential treatment status for a unit facing cutoff c and with score $x \geq c$. In this case, we assume that the functions $D_{di}(x; c)$ are allowed to depend on x only through their first argument, and our notation emphasizes this fact. The observed treatment status in the fuzzy multi-

cutoff RD design is

$$D_i = D_i(X_i, C_i) = D_{0i}(X_i, C_i)\mathbb{I}(X_i < C_i)$$
$$+ D_{1i}(X_i, C_i)1(X_i \geq C_i).$$

Define $D_{0i}(c) := \lim_{x \to c^-} D_{0i}(x, c)$ and $D_{1i}(c) := \lim_{x \to c^+} D_{1i}(x, c)$. Then, for each cutoff $c \in \mathcal{C}$, we can define four subpopulations: local always takers ($D_{1i}(c) = D_{0i}(c) = 1$), local never takers ($D_{1i}(c) = D_{0i}(c) = 0$), local compliers ($D_{1i}(c) > D_{0i}(c)$), and local defiers ($D_{1i}(c) < D_{0i}(c)$).

Within this framework, the smoothness condition (assumption 2 in the sharp multi-cutoff RD setting) can be adapted as follows.

Assumption 7 (Continuity of Regression Functions). For all $c \in \mathcal{C}$: $\mathbb{E}[(Y_{1i}(c) - Y_{0i}(c))D_{1i}(x, c)|X_i = x, C_i = c]$ and $\mathbb{E}[D_{1i}(x, c)|X_i = x, C_i = c]$ are right continuous in x at $x = c$. $\mathbb{E}[(Y_{1i}(c) - Y_{0i}(c))D_{0i}(x, c)| X_i = x, C_i = c]$ and $\mathbb{E}[D_{0i}(x, c)| X_i = x, C_i = c]$ are left continuous in x and $x = c$.

Finally, as it is common in the (causal) instrumental variables literature, we rule out local defiers.

Assumption 8 (Monotonicity). For all $c \in \mathcal{C}$:

$$\mathbb{P}[D_{1i}(c) \geq D_{0i}(c)] = 1.$$

The main identification result for the normalizing-and-pooling approach in the fuzzy multi-cutoff RD design is summarized in the following lemma.
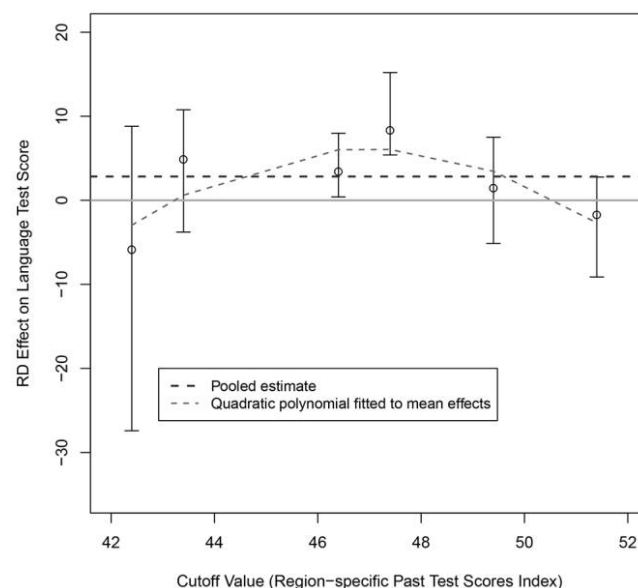

Figure 7. RD effects of P-900 assignment on language test scores

Lemma 2 (Pooled Fuzzy Multi-cutoff RD). If assumptions 2, 3, 6, 7, and 8 hold, the pooled fuzzy RD causal estimand is

$$\tau_{\text{FRD}}^{\text{P}} = \sum_{c \in \mathcal{C}} \mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|D_{1i}(c) > D_{0i}(c),$$
$$X_i = c, \; C_i = c] \, \omega_{\text{FRD}}(c),$$

where

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

This lemma gives an analogue of lemma 1, and can be interpreted in exactly the same way. Furthermore, the same ideas and discussion given in Identification in Multi-Cutoff RD Designs section of the paper, for sharp multi-cutoff RD designs, apply to the fuzzy setting. We do not work through the different assumptions to avoid unnecessary repetition.

RECOMMENDATIONS FOR PRACTICE

We now outline a few simple recommendations for applied researchers. As a starting point, we suggest some visual and descriptive diagnostics to explore the density of observations around each cutoff. If most of the mass in the distribution is near the same cutoff value, then the analyst can treat the design as equivalent to a single-cutoff RD design, since the heterogeneity is minimal. If, in contrast, there are many units exposed to different cutoffs, this simple analysis will reveal that the normalizing-and-pooling approach is combining effects that are heterogeneous in the cutoff value. For multi-cutoff RD designs based on multiparty elections, the analyst should create a histogram of the strongest opponent's vote shares, as we did in figure 2. If the density of this variable is relatively dispersed as in figure 2A then the pooled estimand is potentially heterogeneous. For other types of multi-cutoff RD designs with discrete cutoff variables, researchers can again explore the number of units exposed to each of the cutoff values. When potential heterogeneity is present, the analyst has several options.

First, one could simply pool the estimates and either ignore (i.e., average) the heterogeneity or, alternatively, assume constant treatment effects. Second, one could acknowledge the presence of heterogeneity but leave it unexplored and make the pooled estimand the main object of interest. Third, one could explore whether the pooled estimate is robust to excluding some of the observations. For example, in a case that looks like our Brazil incumbency advantage example, one could split the sample into two subsets: races where the strongest opponent gets 45% or more of the vote, and the rest. If most

of the mass is in the first subset, an interesting question is whether the pooled estimate is actually close to the estimate that uses only this subset. Since the pooled estimand is a weighted average, a low mass of observations below the 45% cut point would receive little weight, but an aberrant treatment effect in this range could lead to an "nonrepresentative" pooled effect.

Next, one could test substantive hypotheses about how the heterogeneity is expected to change from one cutoff to the next and explore these hypotheses and heterogeneity fully, estimating several treatment effects along the cutoff variable. For example, one could formally investigate the presence of monotonic treatment effects along the running variable.

Finally, an important lesson of our framework is that RD designs with multiple cumulative cutoffs are very different from settings in which the cutoffs are noncumulative. In particular, as we discussed, some ignorability assumptions are harder to defend in cumulative multi-cutoff settings. Thus, an important step in the analysis and interpretation of the multi-cutoff RD design is to establish whether the cutoff values are cumulative or noncumulative and evaluate the plausibility of assumptions accordingly.

CONCLUDING REMARKS

The standard RD design assumes that a treatment is assigned on the basis of whether a score exceeds a single cutoff. However, in many empirical RD applications the cutoff varies by unit, and researchers normalize the running variable so that all units face the same cutoff value and a single estimate can be obtained by pooling all observations. This is a useful approach to summarize the average effect across cutoffs, but in some cases it is possible to disaggregate the information contained in the pooled effect and provide a richer description of the underlying heterogeneity in the treatment effect. This heterogeneity can be important from a policy perspective, as it may capture differential RD treatment effects for different values of the running variable.

When there are multiple cutoffs, the pooled RD estimand is a weighted average of the average effects of treatment at every cutoff value, with higher weight given to a particular cutoff value c when there are many units whose scores are close to c. Our formalization of the pooled estimand as a weighted average thus shows that the degree of heterogeneity captured by this estimand will vary on a case-by-case basis depending on the density of the data used in each application. This result continues to be true for both the fuzzy and kink RD designs with multiple cutoffs. We also discussed different assumptions that allow for a causal interpretation of the disaggregated RD treatment effects obtained at different cutoff levels. Importantly, we showed that the plau-

sibility of some of these assumptions depends on whether the cutoffs are cumulative or noncumulative. When cutoffs are cumulative, the cutoff(s) faced by a unit are a deterministic function of the unit's score value; this means that there is a lack of common support of the running variable for units exposed to different cutoffs, making cutoff ignorability assumptions particularly implausible if the running variable is related to the potential outcomes. Thus, a crucial step in the analysis of multi-cutoff RD designs is to determine whether the cutoffs are cumulative.

We also briefly summarized how estimation and inference can be conducted in the multi-cutoff setting and illustrated these steps in the discussion of three empirical examples. Moreover, in the appendix we discuss the connections between our multi-cutoff RD framework and RD designs with multiple scores or running variables. In particular, we show how an RD design with two running variables can be recast as an RD design with one score and multiple cutoffs, a result that highlights the connections between geographic RD designs and the multi-cutoff framework we developed.

Our motivational and empirical examples illustrated the main methodological points of our paper. In the case of Brazilian municipal elections, a multi-cutoff RD design arises because a substantial proportion of Brazil mayoral elections are decided far from the 50% cutoff: in this setting, it is common for the two top parties combined to obtain less than 80% or 70% of the vote. In this scenario, the heterogeneity underlying the pooled estimand can be substantial. As we show for the effect of the PSDB winning on its future electoral victory, the pooled estimate is statistically indistinguishable from zero, but when we use only observations where the vote share obtained by the PSDB's strongest opponent is near 49%, this effect becomes negative and statistically different from zero. In our other two empirical examples, the heterogeneity in the cutoff-by-cutoff RD treatment effects was less clear-cut. For example, our replication of the effects of federal transfers on municipal corruption in Brazil showed that the cutoff-specific effects are broadly consistent with the conclusions from the pooled analysis. Our empirical examples also illustrated in practice the differences between cumulative and noncumulative multi-cutoff RD designs.

In showing that the weights in the pooled approach combine the effects at different cutoffs in a particular way, our framework also suggests that researchers may want to choose different weights relevant to their application. However, as discussed above, the interpretation of this heterogeneity depends on whether the probability that a unit faces a particular cutoff is related to characteristics that correlate with the potential effects of the treatment. Indeed, while identifying the estimands $\tau(c) = \mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|X_i = c, C_i = c]$ for

$c \in \mathcal{C}$ is straightforward, these estimands do not necessarily correspond to the more interesting and policy-relevant estimands $\mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|X_i = c]$ for $c \in \mathcal{C}$ (or, even more difficult, $\mathbb{E}[Y_{1i}(c) - Y_{0i}(c)|X_i = x]$ for $x \neq c \in \mathcal{C}$). In concurrent work, we are investigating different ways to identify this type of estimands in the context of the multi-cutoff RD design, under conditions that allow units to "sort into" the cutoff they face. This is perhaps the most important extension of our work, as our current assumptions effectively treat the variable C_i as exogenous, while in many applications differences in the RD average treatment effect at different values of the cutoff C_i may arise due to "selection" of different unit types into cutoffs.

REFERENCES

Allcott, Hunt. 2015. "Site Selection Bias in Program Evaluation." *Quarterly Journal of Economics* 3 (130): 1117–65.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–55.

Bowers, Jake, Mark M. Fredrickson, and Costas Panagopoulos. 2013. "Reasoning about Interference between Units: A General Framework." *Political Analysis* 21 (1): 97–124.

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

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell. 2016. "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference." Working paper, University of Michigan.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014a. "Robust Data-Driven Inference in the Regression-Discontinuity Design." *Stata Journal* 14 (4): 909–46.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014b. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2015a. "Optimal Data-Driven Regression Discontinuity Plots." *Journal of the American Statistical Association* 110 (512): 1753–69.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2015b. "rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs." *R Journal* 7 (1): 38–51.

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

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

Cattaneo, Matias D., Rocio Titiunik, and Gonzalo Vazquez-Bare. 2016a. "Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality." Working paper, University of Michigan.

Cattaneo, Matias D., Rocio Titiunik, and Gonzalo Vazquez-Bare. 2016b. "rdlocrand: Inference in Regression Discontinuity Designs under Local Randomization." *Stata Journal* (forthcoming).

Caughey, Devin, and Jasjeet S. Sekhon. 2011. "Elections and the Regression Discontinuity Design: Lessons from Close US House Races, 1942–2008." *Political Analysis* 19: 385–408.

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): 1237–58.

de la Cuesta, Brandon, and Kosuke Imai. 2016. "Misunderstandings about the Regression Discontinuity Design in the Study of Close Elections." *Annual Review of Political Science* 19 (forthcoming).

DiNardo, John, and David S. Lee. 2004. "Economic Impacts of New Unionization on Private Sector Employers: 1984–2001." *Quarterly Journal of Economics* 119 (4): 1383–1441.

Dong, Yingying, and Arthur Lewbel. 2015. "Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models." *Review of Economics and Statistics* 97 (5): 1081–92.

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): 259–74.

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (1): 201–9.

Hotz, V. Joseph, Guido Imbens, and Julie H. Mortimer. 2005. "Predicting the Efficacy of Future Training Programs Using Past Experiences at Other Locations." *Journal of Econometrics* 125 (1–2): 241–70.

Keele, Luke J., and Rocío Titiunik. 2015. "Geographic Boundaries as Regression Discontinuities." *Political Analysis* 23 (1):127–55.

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): 675–97.

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): 203–7.

Sinclair, Betsy, Margaret McConnell, and Donald P. Green. 2012. "Detecting Spillover Effects: Design and Analysis of Multilevel Experiments." *American Journal of Political Science* 56 (4): 1055–69.

Skovron, Christopher, and Rocío Titiunik. 2016. "A Practical Guide to Regression Discontinuity Designs in Political Science." Working paper, University of Michigan.

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): 107–41.