CAN CLOSE ELECTION REGRESSION DISCONTINUITY DESIGNS IDENTIFY EFFECTS OF WINNING POLITICIAN CHARACTERISTICS?[*]

JOHN MARSHALL[†]

MAY 2022

Politician characteristic regression discontinuity (PCRD) designs leveraging close elections are widely used to isolate effects of an elected politician characteristic on downstream outcomes. Unlike standard regression discontinuity designs, treatment is defined by a predetermined characteristic that could affect a politician's victory margin. I prove that, by conditioning on politicians who win close elections, PCRD estimators identify the effect of the specific characteristic of interest *and all compensating differentials*—candidate-level characteristics that ensure elections remain close between candidates that differ in the characteristic of interest. Avoiding this asymptotic bias generally requires assuming either that the characteristic of interest does not affect candidate vote shares *or* that no compensating differential affects the outcome. Since theories of voting behavior suggest that neither strong assumption usually holds, I further analyze the implications for interpreting continuity tests and consider if and how covariate adjustment, bounding, and recharacterizing treatment can mitigate the post-treatment bias afflicting PCRD designs.

[†]Department of Political Science, Columbia University. Email: jm4401@columbia.edu.

1   Introduction

Regression discontinuity (RD) designs have become a staple of the quantitative social scientist's methodological toolkit. RD designs leverage treatment assignments that change discontinuously at a known threshold in a forcing variable to identify treatment effects for observations around that threshold (see Cattaneo and Titiunik forthcoming). While external validity can be limited, such designs are often regarded as the observational method best with greatest internal validity. As researchers have increasingly focused on estimating causal effects, the use of RD designs in political science has exploded over the last decade (de la Cuesta and Imai 2016).

A particularly popular version of the RD design uses close elections to estimate effects of a specific characteristic of elected politicians on downstream electoral, policy, and constituent outcomes. I will call this application a politician characteristic regression discontinuity (PCRD) design. Studies from across the globe have used PCRD designs to compare narrowly-elected politicians who differ in terms of a given predetermined characteristic, usually with the objective of holding observable and unobservable confounders constant.[1] Appendix Table A1 lists 126 published articles—often in prestigious journals—that estimate downstream effects of ascriptive characteristics (gender, race or ethnicity, clan, religious identity), prior actions of politicians (criminal history, prior incumbency, seniority), labels politicians sort into (party membership, ideology), and institutional status (partisan alignment with other levels of government, term limit status). Indeed, PCRD designs appear to facilitate opportunities to study how electoral selection affects representation, accountability, and participation that are only limited by a researcher's capacity to measure politician characteristics of interest.[2]

Although the appeal of credibly estimating effects of winning candidate characteristics is obvious, whether PCRD designs can isolate the effect of a specific *X*—the characteristic, or bundle of characteristics, of interest—among politicians in close elections is not. Indeed, this article demon-

---

[1]The estimand is rarely stated explicitly. Most studies imply that PCRD designs identify "all else equal" effects of a particular characteristic. The same approach has been used for primary elections.

[2]While potentially relevant, multiple testing is not this article's primary focus.

strates that a non-standard feature of this application of the RD design causes PCRD designs to generally identify compound treatment effects, rather than the local average treatment effect (LATE) of X. I will show that this confounding can only be avoided by imposing strong—and often implausible or unverifiable—additional assumptions.

The source of bias emanates from the difference between standard RD and PCRD designs. Standard RD designs define treatment as falling above or below a threshold. For example, close elections have been used to vary whether a candidate or party was *elected or not* to estimate financial returns to holding office (e.g. Eggers and Hainmueller 2009) and incumbent party electoral advantages (e.g. Lee, Moretti and Butler 2004). In contrast, the treatment in PCRD designs—which instead seek to estimate the LATE of an elected politician characteristic—is defined by possessing (or not) predetermined characteristic X, *conditional on narrowly winning an election*. Beyond targeting different estimands, the mechanics of PCRD designs differ from standard RD designs in two important ways. First, as Sekhon and Titiunik (2012) have noted, close elections do not as-if randomly assign characteristic X. The potential correlation between X and other politician characteristics creates the risk of confounding or necessitates reinterpreting the estimand as a compound treatment. Second, and more subtly, restricting attention to close elections entails conditioning on candidate vote shares that may be affected by X. As I show below, the former difference is fairly frequently acknowledged by researchers, but the latter is largely unrecognized. This article focuses on the second issue, showing that PCRD designs generally introduce bias—even when X is independent of other predetermined variables and the weak continuity assumption underpinning standard RD designs holds.

By expressing PCRD designs within the continuity framework of standard RD designs, I show that conditioning on close elections between candidates that differ in terms of characteristic X causes PCRD estimators to identify the LATE of electing a candidate of type *X combined with a (differential-weighted) LATE of any compensating differentials*. While PCRD designs ensure continuity across the districts different types of candidates are elected from, the vector of compensating differentials $\mathbf{Z}$ which generates this (asymptotic) bias is defined by characteristics of individual

candidates that: (a) the researcher regards as theoretically distinct from X; and (b) ensure candidates of type X remain in close elections with candidates not of type X. For example, in seeking to isolate the effect of gender, competence would be a compensating differential if women in close elections were more competent than men in close races because voters were biased against women.

My main identification result establishes that, even when X is (unconditionally) independent of $\mathbf{Z}$, PCRD designs require strong additional assumptions to isolate the effect solely attributable to characteristic X. Specifically, identification requires that—at the discontinuity—either: (i) X does not affect the winning candidate's victory margin; or (ii) no compensating differential in $\mathbf{Z}$ affects the outcome of interest Y. These assumptions are analogous to the conditions under which the bias associated with conditioning the sample on a post-treatment variable disappears (see Hernán and Robins 2011). Where neither additional condition holds, compensating differentials cause PCRD designs to underestimate (overestimate) the LATE of X when the net effect of $\mathbf{Z}$ affects the candidate's vote share and the outcome Y in the same (opposite) direction.

I highlight three implications for applied research. First, to claim that PCRD estimates can isolate the effect of X by design, researchers should explicitly state and support one of the additional assumptions just described. However, these assumptions are difficult to empirically substantiate and are theoretically implausible when voters observe X and believe it will affect outcomes they care about. Second, in the likely event that neither assumption can be sustained, strategies for mitigating threats to internal validity vary in their effectiveness. Whereas PCRD estimates that reject a null hypothesis could be combined with candidate-level (dis)continuity tests to bound the LATE of X, covariate adjustment strategies cannot generally prevent biases induced by post-treatment conditioning. Indeed, PCRD designs do not imply that candidate-level covariates should be continuous at the point of discontinuity. Third, researchers might consider reinterpreting PCRD estimates as capturing (weighted) effects of X and $\mathbf{Z}$. Where plausible compensating differentials can be measured, candidate-level discontinuity tests can now help to interpret this compound treatment. However, by focusing on a less well-defined and possibly heterogeneous conception of treatment, researchers cannot isolate the effect of X—which is often desirable from a theoretical or policy perspective.

By clarifying the interpretation of PCRD designs, this article makes several contributions. First, I provide the first systematic account of the challenges that arise when RD designs define treatment by a variable that can also affect the forcing variable. While some articles have noted study-specific issues relating to how conditioning on close elections introduces compensating differentials (e.g. Gagliarducci and Paserman 2012), 118 of 126 published studies using PCRD designs do not even loosely acknowledge the possibility of compensating differentials. Second, I move beyond proving the inconsistency of PCRD estimators by identifying strong additional assumptions under which the LATE of X can be isolated, establishing when PCRD designs underestimate and overestimate the LATE of X, and evaluating strategies to mitigate bias and reinterpret PCRD estimates. Third, this article illustrates the need to understand why candidates end up in close elections, and thus complements recent work emphasizing the theoretical implications of empirical models (e.g. Ashworth, Berry and Bueno de Mesquita 2021; Bueno de Mesquita and Tyson 2020; Eggers 2017) and the large econometric literature documenting how sample selection generates bias (e.g. Heckman 1979).

The post-treatment bias this article highlights differs from prior critiques of RD designs leveraging close elections. Extant studies have examined other ways through which compound treatments can confound causal attribution, including where multiple treatments are assigned at the same threshold (Eggers et al. 2018), where correlated characteristics—like Black politicians in the U.S. overwhelmingly being Democrats—are bundled together (e.g. Bucchianeri 2018; Ferreira and Gyourko 2014; Hall 2015), and where treatment affects downstream behaviors such as future candidacy decisions (Eggers 2017; Sekhon and Titiunik 2012). Researchers have also debated whether election outcomes are determined by chance at the discontinuity (Caughey and Sekhon 2011; Eggers et al. 2015) and highlighted the sensitivity of RD estimates and inference to bandwidth sizes and functional form assumptions (Cattaneo and Titiunik forthcoming; Gelman and Imbens 2019). However, the conceptual problems raised by this article still arise when the standard RD continuity assumption holds and enough data exists for consistent estimation of conditional expectations at the threshold.

2 PCRD designs in practice

This section builds intuition for the issues that arise when PCRD designs are used to isolate effects of elected politician characteristics. I first describe the design and its potential problems through the lens of two commonly-studied characteristics—gender and party affiliation. I then review published articles to characterize how PCRD designs have been used and summarize the identification concerns they address.

2.1 Electing women

In mt first example, researchers compare outcomes across polities where women and men were elected. Extant studies have used PCRD designs to estimate effects of electing women on policy priorities (e.g. Clots-Figueras 2011; Ferreira and Gyourko 2014), turnout among women and women running for office at future elections (e.g. Broockman 2014; Ferreira and Gyourko 2014), and government instability (Gagliarducci and Paserman 2012). Others have similarly examined the effect of women winning primary elections on general election results (Bucchianeri 2018). Many of these studies define their estimand as the effect of electing a woman instead of a man, often implicitly holding other characteristics of the candidate constant. For example, Ferreira and Gyourko (2014:24) describe their design as estimating the "effect of gender" and Clots-Figueras (2011:665) describes her design as estimating "the effect of a legislator's gender."

Before illustrating the identification problem at hand, it should be emphasized that defining gender as a treatment that is conceptually distinct from other candidate characteristics is challenging. This is because gender is often viewed as an inherently bundled treatment comprising various correlated features;[3] women who win close elections may espouse different policies, possess different qualifications, or have different personalities from men who win close elections. To isolate the effect of electing women, a researcher must distinguish the bundle of characteristics that differentiate women and men candidates—the definition of treatment—from the characteristics they regard

---

[3]I abstract from whether ascriptive characteristics are manipulable.

as distinct from gender—the potential confounders. Gender is a particularly challenging example, but the need to explicitly define treatment applies equally to other characteristics—such as prior experience or partisan alignment with other relevant politicians—that may be easier to distinguish conceptually from potential confounders.

PCRD designs typically then estimate the effect of politician gender in single-member plurality races by comparing outcomes in "treated" districts where a woman was just elected in a race against a man with outcomes in "control" districts where a man was just elected in a race against a woman. Invoking the standard RD assumption of continuity in potential outcomes or local randomization (see Cattaneo and Titiunik forthcoming), it is usually argued that the two types of districts will, in expectation, be identical in terms of district-level covariates and all other individual-level characteristics of winning candidates at the point of discontinuity. If this were the case, only gender would change at the discontinuity and PCRD designs would identify the effect of electing a woman over an otherwise-similar man. To ease exposition, I label the set of potential confounders—the bundle of characteristics that are conceptually distinct from gender—as "competence" and assume that voters prefer more competent candidates because they achieve positive district outcomes—employment in this example.

Using black to indicate the data researchers can observe, Figure 1a plots hypothetical conditional expectations of district employment as a function of a woman's victory margin Δ_d. Cases to the right of the vertical line, where a woman is elected ahead of a man ($X_d = 1$), are treated. The relevant counterfactual for isolating the effect of gender, shown in gray, is *equally competent* men who win elections against women. The difference in employment at the discontinuity between districts where men and women were elected, τ_{PCRD}, is the LATE of electing a woman over a man in a close election.

Is it reasonable to assume *at the candidate level* that men who narrowly win are equally competent as women who narrowly win? Suppose that—holding competence fixed—voters are more likely to vote for men (e.g. Lawless 2015); this could arise from stereotyping, media attention, or differential support from political elites. To be in close races with women, men must then possess

Figure 1: Hypothetical examples of PCRD design biases

(a) Gender

(b) Partisan affiliation
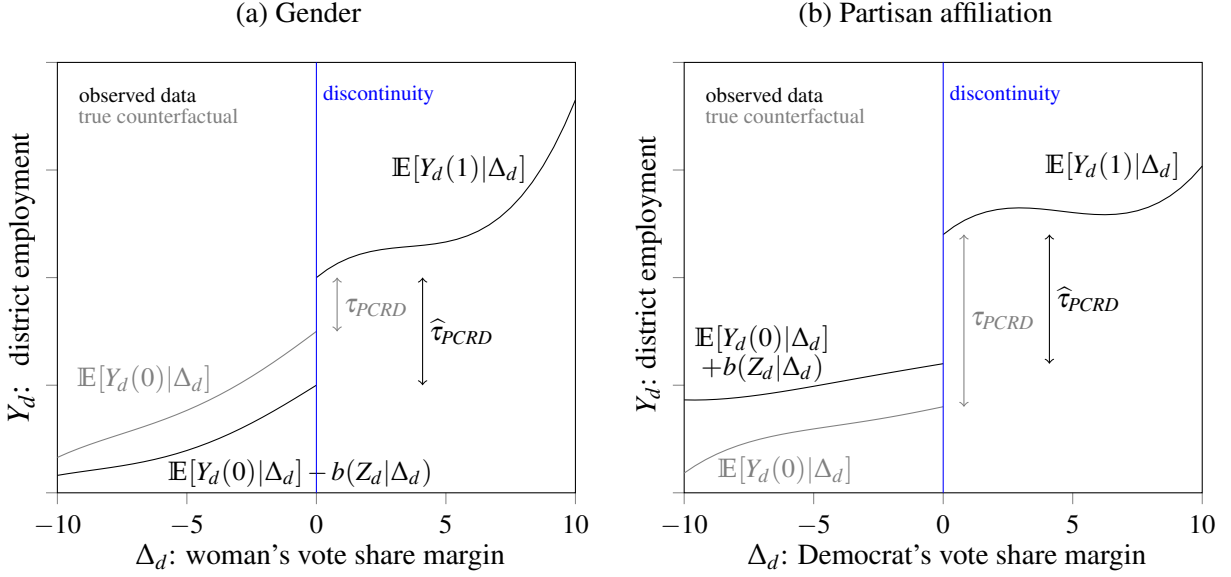
Notes: For a given district d, $Y_d(X_d)$ denotes the potential outcome, $X_d \in \{0,1\}$ denotes the characteristic of interest of the winning candidate, Δ_d denotes the winning margin of the candidate that possesses characteristic X, Z_d denotes the compensating differential of a winning candidate, and the function b captures the effect of Z_d on the district outcome. τ_{PCRD} denotes the true LATE of X_d and $\widehat{\tau}_{PCRD}$ denotes a consistent estimate of the difference between the limits at $\Delta_d = 0$ captured by a typical PCRD estimator.

lower levels of competence in expectation than the women against whom they competed. The observed conditional expectation function for the men who won is shown in black to the left of the vertical line, where $b(Z_d|\Delta_d)$ denotes the reduction in employment due to elected men possessing less competence Z_d than elected women at a given vote margin. Even when each function in black is consistently estimated at $\Delta_d = 0$, the PCRD estimate $\widehat{\tau}_{PCRD}$ is upwardly (asymptotically) biased in this example because it is confounded by competence—the compensating differential required for women to be in close races with men when voters are biased against women.

2.2 Electing politicians from different parties

Analogous challenges apply to estimating differences due to a candidate's party affiliation. Studies using PCRD designs compare outcomes between political units that elected candidates from different parties (e.g. Gerber and Hopkins 2011; Lee, Moretti and Butler 2004; Pettersson-Lidbom 2008). These studies describe the design as capturing the "effect of a Democratic victory" (Ger-

ber and Hopkins 2011:335) or "causal estimates of the effect of party control" (Pettersson-Lidbom 2008:1037). This example focuses on the hypothetical effect of electing Democrats over Republicans on district employment.

The researcher must again specify what does and does not constitute part of the partisan affiliation treatment. To illustrate, I assume that party affiliation captures a common set of policy positions or ideology and that Democrats are more popular with voters on average. I again consider candidate "competence" as the compensating differential, and assume that more competent politicians increase employment. Other potential compensating differentials could include alignment with higher-level incumbents or prior performance in office.

PCRD designs then compare districts where Democrats and Republicans barely won close elections against candidates from the other party. Since Democrat candidates are more popular than Republican candidates in this example, Republican candidates need to be more competent to counteract this disadvantage. By conditioning on close elections, PCRD designs then compare relatively incompetent Democrats with relatively competent Republicans. As Figure 1b illustrates, the PCRD estimate $\widehat{\tau}_{PCRD}$ understates the effect of being a Democrat in this example because narrowly-elected Democrats possess less competence Z_d than narrowly-elected Republicans.

2.3 Limited awareness of compensating differentials

To examine awareness of these potential issues, I used Google Scholar to identify 126 published articles employing PCRD designs.[4] The earliest article was published in 2004, but 78% of articles have been published since 2015. While 38% of studies have focused on executive and legislative elections in the U.S., PCRD designs are also commonly applied to the election of individual politicians or changes in legislative representation or control in majoritarian and proportional representation systems in Asia, Europe, and South America. These articles have consistently appeared in prominent journals in political science and economics: 21% were published in the *American Journal of Political Science*, *American Political Science Review*, or *Journal of Politics*, while 5% were

---

[4]These are listed in Appendix Table A1.

published in the *American Economic Review*, *Econometrica*, the *Quarterly Journal of Economics*, or the *Review of Economic Studies*. According to Google Scholar, these studies had collectively amassed 11,774 citations by March 8, 2022.

After reading each article, I hand-coded whether the article demonstrated awareness of four potential threats to internal validity. Specifically, I coded whether an article: (i) assessed continuity in potential outcomes by testing for discontinuities in district- or candidate-level covariates; (ii) assessed the same continuity assumption using density tests to test for sorting around the discontinuity; (iii) recognized that candidate characteristics may come as bundles due to unconditional correlations between characteristics; and (iv) discussed the risk of inducing or altering correlations between candidate characteristics by conditioning on close elections. The first two threats are benchmarks that apply to all RD designs (see Cattaneo and Titiunik forthcoming), the third could apply to other RD designs but is especially relevant for PCRD designs, and the fourth is specific to PCRD designs. Even brief and suggestive references to an issue were coded positively.

The results in Table 1 indicate that applied researchers are already aware of the importance of validating the continuity assumption and, to a lesser degree, that candidate characteristics come as bundles that are hard to separate. Panel A shows that 91% of articles conducted balance or continuity tests and 75% conducted density tests like the one proposed by McCrary (2008). Panel B shows that both strategies for validating the continuity assumption have become more prevalent over time. Furthermore, 33% of articles demonstrated awareness of the possibility that the candidate characteristic of interest might be unconditionally correlated with other characteristics. As panel D demonstrates, few studies acknowledged that candidates from different parties may also differ in terms of other candidate characteristics. However, the majority of studies seeking to estimate effects of candidate education, gender, ideology, incumbency status, and vocation discussed this issue.

In contrast, very few studies demonstrated any awareness of the issue motivating this article—the risk of inducing or altering correlations between the candidate characteristic of interest and compensating differentials by conditioning on close elections. Indeed, only 6% of articles even

Table 1: PCRD designs demonstrating awareness of different threats to internal validity

	Number of articles	...of which, employ/demonstrate awareness of...			
		Covariate continuity tests	Sorting/ density tests	Correlated characteristics	Compensating differentials
Panel A: All articles					
All articles	126	115	94	42	8
Panel B: Articles by five-year period					
2002-2006	1	1	0	0	0
2007-2011	10	6	1	3	0
2012-2016	39	33	29	15	5
2017-forthcoming	76	75	64	24	3
Panel C: Articles by region					
Africa	0	0	0	0	0
Asia	21	20	19	7	1
Europe	26	24	24	8	2
Middle East	4	4	4	1	0
North America	50	44	27	13	3
Oceania	0	0	0	0	0
South America	24	22	19	13	2
Cross-continental	1	1	1	0	0
Panel D: Articles by politician characteristic					
Partisan alignment across tiers of government	17	15	14	4	1
Criminal history	4	4	4	1	0
Education	4	3	2	2	1
Gender	24	23	18	17	5
Ideology	2	2	1	2	0
Incumbency, term limit status, or seniority	11	9	7	8	2
Partisan affiliation	58	54	42	5	0
Race, ethnicity, religion, or clan	7	6	6	4	0
Pre-office vocation	5	5	4	3	0
Panel E: Articles by type of electoral discontinuity					
Individual politician (executive or legislator)	101	92	73	36	8
Legislature majority	13	11	11	1	0
Legislature seat share	13	13	10	5	0
Party representation threshold	2	2	2	0	0

Notes: Data are based on the author's hand-coded classifications. Some articles fall into multiple categories.

loosely mention this issue. These articles usually note that elections could be close because a difference in the characteristic of interest is counterbalanced by differences in terms of other characteristics; in 5 of the 8 cases, this arises from the specific concern that women in close races may differ from men due to voter biases or campaign disadvantages. However, the origins and implications of such compensating differentials received limited discussion and were often quickly dismissed, despite the fact that compensating differentials between candidates will generally exist

in PCRD designs—as this article demonstrates.

One article delves deeper into this issue. Gagliarducci and Paserman (2012:1031) note that identifying the effect of elected women using a PCRD design requires that "the vote share of each candidate must not depend directly on gender." They argue that this assumption is plausible in their study examining the effect of women narrowly elected as mayors on early municipal government termination in Italy, if voters do not select candidates to maintain government stability, are unaware that mayor gender affects government stability, or only select candidates on the basis of factors unrelated to gender that could affect government stability. This article generalizes the conditions under which PCRD designs identify the desired estimand, characterizes the nature of asymptotic bias when these strong conditions do not hold, and discusses bias mitigation strategies and alternative conceptualizations of treatment.

3 Theoretical analysis

This section first recaps how standard RD designs work in the context of plurality elections in single member districts, before explaining how PCRD designs differ.[5] I then show how a post-treatment bias introduced by these differences prevents PCRD designs from isolating the effect of the characteristic of interest in a stylized example. I finally provide general results demonstrating that additional assumptions—which are far stronger than the standard RD continuity assumption—are required to identify the effects often attributed to PCRD designs by applied researchers.

3.1 Standard RD designs

In the close election application of RD designs, each candidate i in district d receives share $V_{id} \in [0, 1]$ of the votes cast between the top two candidates. The continuously-distributed forcing variable is the difference between V_{id} and the vote share V_{jd} of the other most popular candidate $j \neq i$

---

[5]Designs adapted to legislative chambers (Clots-Figueras 2011), proportional representation elections (Folke 2014), and to leverage discontinuities in control of legislative bodies (Pettersson-Lidbom 2008) differ in some respects. However, similar challenges apply to isolating effects of elected politician characteristics.

in the district: $\Delta_{id} := V_{id} - V_{jd} \in [-1, 1]$. The following treatment variable then indicates whether candidate i won the election in district d or not:[6]

$$T_{id} := \begin{cases} 1 & \text{if } \Delta_{id} > 0 \\ 0 & \text{if } \Delta_{id} \leq 0. \end{cases} \tag{1}$$

In addition to observing T_{id} based on which candidate wins the race, researchers also observe an outcome variable Y_{id} for each candidate. The potential outcome $Y_{id}(T_{id}) \in \mathbb{R}$ depends on whether a candidate wins office. This representation encodes the SUTVA assumption that i's potential outcomes are not affected by the treatment status of other candidates and that there is a single version of treatment. Since only one potential outcome can be observed, the observed outcome is related to potential outcomes by $Y_{id} = T_{id}Y_{id}(1) + (1 - T_{id})Y_{id}(0)$.

The standard RD design requires the following weak continuity assumption:[7]

Assumption 1. *Potential outcomes $Y_{id}(T_{id})$ satisfy:*

(a) *Continuity from above:* $\lim_{v \downarrow 0} \mathbb{E}[Y_{id}(1)|\Delta_{id} = v] = \mathbb{E}[Y_{id}(1)|\Delta_{id} = 0]$;

(b) *Continuity from below:* $\lim_{v \uparrow 0} \mathbb{E}[Y_{id}(0)|\Delta_{id} = v] = \mathbb{E}[Y_{id}(0)|\Delta_{id} = 0]$.

This assumption states that, at the point of discontinuity, potential outcomes do not vary discontinuously in any way other than whether a given candidate won the election. In the case of close elections, this is plausible because factors exogenous to candidate characteristics, such as election day weather, generate random variation in which candidate wins; Eggers et al. (2015) provide evidence supporting this claim from ten countries across the world.

When continuity holds, the LATE of T_{id} at the point of discontinuity—denoted by $\tau_{RD} := \mathbb{E}[Y_{id}(1) - Y_{id}(0)|\Delta_{id} = 0]$—can be identified by comparing observed outcomes between candidates that narrowly won and narrowly lost. Applied researchers typically employ an RD estimator

---

[6]For simplicity, I assume that i does not win if $V_{id} = V_{jd}$.

[7]The local randomization approach imposes stronger assumptions (Cattaneo and Titiunik forthcoming).

of the following form:

$$\widehat{\tau}_{RD} = \widehat{\mu}_+(0) - \widehat{\mu}_-(0), \tag{2}$$

where $\widehat{\mu}_+(s)$ and $\widehat{\mu}_-(s)$ are estimators of $\lim_{v \downarrow s} \mathbb{E}[Y_{id}|\Delta_{id} = v]$ and $\lim_{v \uparrow s} \mathbb{E}[Y_{id}|\Delta_{id} = v]$, respectively. The state of the art involves estimating $\widehat{\mu}_+(0)$ and $\widehat{\mu}_-(0)$ using local polynomial regressions either side of the discontinuity and correcting for a consistent estimate of the misspecification bias that arises from approximating the unknown functional form of $\mathbb{E}[Y_{id}|\Delta_{id} = v]$ (Calonico, Cattaneo and Titiunik 2014; Calonico et al. 2019). To trade off the finite sample biases and precision gains of including observations further from the discontinuity, researchers often use a data-driven procedure to select the bandwidth that minimizes the mean squared error of $\widehat{\tau}_{RD}$. See Cattaneo and Titiunik (forthcoming) for an excellent review of RD estimation and inference methods.

To focus on the asymptotic bias—the bias as the sample size tends to infinity—that arises with PCRD designs, this article abstracts from estimation challenges. Specifically, for a random sample of n elections drawn from a large population, I assume that:

Assumption 2. *For any conditioning set W, $\widehat{\mu}_+(0|W)$ and $\widehat{\mu}_-(0|W)$ are consistent estimators with bounded variance.*

Following Hahn, Todd and Van der Klaauw (2001) and Imbens and Lemieux (2008), it is well-established that:

Proposition 1. *Under Assumptions 1 and 2, $\widehat{\tau}_{RD}$ is a consistent and asymptotically unbiased estimator of τ_{RD}.*

Proof: see Appendix for all proofs. ∎

Within political science, this type of RD design has proved popular for estimating the consequences of *being elected*. One strand of this literature has explored the effect of being elected to office on a candidate's downstream wealth (e.g. Eggers and Hainmueller 2009). Another strand has studied the effect of winning elections on subsequent election outcomes. Since the decision to

14

run in future elections often depends on winning a close election, researchers interested in isolating the effect of winning office, conditional on running, on future electoral success have focused on the electoral outcomes of parties linked to the winning and losing candidates where parties always run for office. This has spawned a large literature measuring party incumbency advantages. In the U.S., narrowly-elected incumbent parties are substantially more likely to win future elections in the same district (e.g. Lee, Moretti and Butler 2004). These studies are united by comparing downstream outcomes *between election winners and losers*.

3.2 How PCRD designs differ from standard RD designs

Whereas close election applications of standard RD designs estimate effects of a politician getting elected, PCRD designs instead seek to isolate effects of a specific characteristic of elected politicians on downstream outcomes. Rather than compare winning and losing candidates, PCRD designs compare politicians who all narrowly won elections in separate districts but differ according to a predetermined binary characteristic denoted by $X_{id} \in \{0,1\}$.[8] Characteristics of empirical interest have included gender, race, vocational experience, criminality, prior incumbency, partisan affiliation, and partisan alignment.

Isolating the effect of a characteristic of interest is difficult because politicians are defined by many characteristics that tend to be correlated. For example, elected women in the U.S. are more likely to be Democrats, politicians who have engaged in corruption are more likely to be aligned with higher-level politicians, or politicians from traditional parties are more likely to be experienced. If the characteristic—or bundle of characteristics—of interest X_{id} is correlated with a vector of K distinct candidate-level characteristics $\mathbf{Z}_{id} \in \mathbb{R}^K$, any effect of X_{id} may be confounded by the effects of $\mathbf{Z}_{id}$. To define their target estimand, researchers must decide which characteristics should and should not be included in their treatment of interest; put differently, they must decide which characteristics are conceptually distinct from the characteristic of interest. This is an inexact science. Some researchers (explicitly or implicitly) assert a characteristic of interest and seek to

---

[8]Non-binary characteristics could compare any two characteristic discrete values or bins.

substantiate the claim that only this characteristic drives any effect. Others define their treatment as a bundle of correlated characteristics that are distinct from some other characteristics.[9]

I abstract from the challenge of interpreting correlated characteristics, which is already acknowledged in a third of articles reviewed in Table 1. Rather, I show that PCRD designs introduce a form of post-treatment bias even when characteristic X_{id} is (unconditionally) independent of other relevant characteristics, e.g. if X_{id} was randomly assigned. Accordingly, I will at times assume that:

Assumption 3. *Characteristic X_{id} is independent of $\mathbf{Z}_{id}$ and $\mathbf{Z}_{jd}$: $X_{id} \perp\!\!\!\perp \mathbf{Z}_{id}, \mathbf{Z}_{jd}$.*

This assumption will clarify that biases emerge in PCRD designs even in a "best case scenario" where—at least among politicians who could end up in close races—X_{id} is independent of i's conceptually-distinct characteristics $\mathbf{Z}_{id}$ and the conceptually-distinct characteristics $\mathbf{Z}_{jd}$ of their chief competitor j. However, Assumption 3 is relaxed for more general theoretical results.

In shifting attention to the *type* of politician who wins, the unit of analysis in PCRD designs is the district. The district-level forcing variable is then $\Delta_d := V_{1d} - V_{0d} \in [-1,1]$, where V_{1d} and V_{0d} respectively denote the vote shares of the most popular politician of type $X_{id} = 1$ and $X_{id} = 0$ in district d. Districts where the top two candidates are of the same type are excluded. The corresponding district-level treatment indicates whether a candidate of type $X_{id} = 1$ won the election:

$$X_d := \begin{cases} 1 & \text{if } \Delta_d > 0 \\ 0 & \text{if } \Delta_d \leq 0. \end{cases} \tag{3}$$

District-level potential outcomes depend on X_d, which reflects the individual-level potential outcomes of the type of politician who won: $Y_d(X_d) = X_d Y_{1d}(1) + (1 - X_d)Y_{0d}(1)$. For example, $Y_d(1)$ and $Y_d(0)$ could correspond to the district-level outcome when the elected candidate was a woman and a man, respectively. In PCRD designs comparing observations of Y_d across districts, the politician-level potential outcome $Y_{id}(0)$ is neither relevant nor well-defined because losing

---

[9]I analyze the case where researchers view *all* characteristics as a single bundle in Section 4.3.

politicians do not enter office.

The LATE of interest in PCRD designs is the difference in potential outcomes across districts with close elections where politicians of different types were elected: $\tau_{PCRD} := \mathbb{E}[Y_d(1) - Y_d(0)|\Delta_d = 0]$. This is typically estimated using the following PCRD estimator:

$$
\begin{aligned}
\widehat{\tau}_{PCRD} &= \lim_{v\downarrow 0}\mathbb{E}\widehat{[Y_d|\Delta_d = v]} - \lim_{v\uparrow 0}\mathbb{E}\widehat{[Y_d|\Delta_d = v]} \\
&= \widehat{\mu}_+(0|X_{id} = 1, X_{jd} = 0) - \widehat{\mu}_+(0|X_{id} = 0, X_{jd} = 1).
\end{aligned}
\tag{4}
$$

The second line rewrites district-level outcomes Y_d in terms of candidate-level outcomes Y_{id} to distinguish standard RD designs from PCRD designs: whereas standard RD designs compare candidates either side of the threshold in the forcing variable, PCRD designs compare narrow winners on one side of the threshold who differ in terms of X_{id}. Consequently, this non-standard RD design conditions on a predetermined difference between X_{id} and X_{jd} that could also affect the forcing variable Δ_d. I next show how this distinction prevents $\widehat{\tau}_{PCRD}$ from consistently estimating τ_{PCRD}, even when Assumptions 1-3 hold.

3.3 Bias in PCRD designs with a single compensating differential

To build intuition, I start with a simple case where a single compensating differential $Z_{id} - Z_{jd}$ ensures that the race between candidates i and j in district d is close despite only one candidate possessing characteristic X_{id}. Let characteristic X_{id} help candidate i win votes, e.g. by being the incumbent, representing a popular party, or not suffering gender-based biases. The compensating differential will offset this advantage, e.g. because the candidate for whom $X_{id} = 1$ is less competent, more malfeasant, or less politically connected than the candidate for whom $X_{id} = 0$. In addition to affecting candidate vote shares, the compensating differential $Z_{id} - Z_{jd}$ will also affect district-level outcomes that depend on the winning candidate's level of Z_{id}.

My stylized example captures these roles of X_{id} and $Z_{id} - Z_{jd}$ in the following functional forms:

$$V_{id} = \alpha \frac{X_{id} - X_{jd}}{2} + \beta \frac{Z_{id} - Z_{jd}}{2} + \frac{\varepsilon_{id} - \varepsilon_{jd}}{2}, \tag{5}$$

$$Y_{id}(1) = \tau X_{id} + \gamma Z_{id} + \upsilon_d, \tag{6}$$

where $\alpha \geq 0$ and $\beta > 0$ imply that possessing more of characteristic X_{id} or Z_{id} increases a candidate's vote share,[10] while τ and γ respectively denote (constant) effects of X_{id} and Z_{id} on district-level outcomes. I further assume that $Z_{id} - Z_{jd} \sim N(0, \sigma_Z^2)$ and $\varepsilon_{id} - \varepsilon_{jd} \sim N(0, \sigma_\varepsilon^2)$ are normally distributed and drawn independently of both X_{id} and each other,[11] while υ_d is district-level noise that is drawn independently of all other variables. By embedding Assumption 3 in these distributional assumptions, any bias in $\widehat{\tau}_{PCRD}$ must emerge from the PCRD design.

3.3.1 Derivation of asymptotic bias

Although X_{id} is independent of $Z_{id} - Z_{jd}$, PCRD designs can generate a correlation by conditioning on close elections where two narrow winners with different characteristics obtain similar vote shares. To see why, note that the limiting case of close elections—where candidates within a given district receive equal numbers of votes—implies:

$$\Delta_d = \alpha + \beta(Z_{1d} - Z_{0d}) + \varepsilon_{1d} - \varepsilon_{0d} = 0, \tag{7}$$

where the electorally-advantaged candidate of type $X_{id} = 1$ is denoted by $i = 1$ and the candidate of type $X_{id} = 0$ is denoted by $i = 0$. A tie between these candidate types occurs because there is a countervailing compensating differential ($Z_{1d} < Z_{0d}$) and/or because candidate 1 encountered unfortunate random shocks ($\varepsilon_{1d} < \varepsilon_{0d}$).

---

[10]α and β are positive for simplicity, but need not be restricted. These coefficients as common across candidates because X_{id} and Z_{id} characterize differences between candidates.

[11]$V_{id} \in [0,1]$ can be violated with normal distributions. However, the general results do not impose unbounded distributions, while $V_{id} \in [0,1]$ almost always holds when σ_Z^2 and σ_ε^2 are small.

The PCRD estimator then converges to the following quantity:

$$\widehat{\tau}_{PCRD} \xrightarrow{p} \lim_{v\downarrow 0}\mathbb{E}[Y_{id}|\Delta_{id}=v,X_{id}=1,X_{jd}=0] - \lim_{v\downarrow 0}\mathbb{E}[Y_{id}|\Delta_{id}=v,X_{id}=0,X_{jd}=1]$$

$$= \mathbb{E}[Y_{id}(1)|\Delta_{id}=0,X_{id}=1,X_{jd}=0] - \mathbb{E}[Y_{id}(1)|\Delta_{id}=0,X_{id}=0,X_{jd}=1]$$

$$= \tau + \gamma\mathbb{E}[Z_{1d}-Z_{0d}|\alpha+\beta(Z_{1d}-Z_{0d})+\varepsilon_{1d}-\varepsilon_{0d}=0]$$

$$+\mathbb{E}[\upsilon_d|\Delta_{id}=0,X_{id}=1,X_{jd}=0] - \mathbb{E}[\upsilon_d|\Delta_{id}=0,X_{id}=0,X_{jd}=1]$$

$$= \tau \underbrace{-\gamma\frac{\alpha\beta\frac{\sigma_Z^2}{\sigma_\varepsilon^2}}{1+\beta^2\frac{\sigma_Z^2}{\sigma_\varepsilon^2}}}_{\substack{\text{asymptotic bias relative} \\ \text{to the LATE of } X_{id}}}, \tag{8}$$

where the first line follows from consistent estimation of conditional expectations (Assumption 2), the second line follows from continuity (Assumption 1(a)) and the district-level outcome being the potential outcome of the winning candidate in that district, the third line substitutes the functional forms for potential outcomes and vote shares, and the final line applies the distributional assumptions on $Z_{id}-Z_{jd}$, $\varepsilon_{id}-\varepsilon_{jd}$, and υ_d.[12] Since PCRD designs compare candidates that win elections, only part (a) of Assumption 1 is needed. Although this weakens the standard RD continuity assumption, it is difficult to imagine contexts where part (a) held but part (b) did not.

This derivation shows that PCRD designs can induce a form of post-treatment bias which yields inconsistent estimates of the effect of characteristic X_{id}. Where X_{id} affects V_{id}, this is because the event $\Delta_{id}=0$ upon which the LATE is conditioned is itself affected by X_{id}. Satisfying $\Delta_{id}=0$, and thus $\Delta_d=0$, when $X_{id}\neq X_{jd}$ generally requires the existence of a compensating differential $Z_{id}\neq Z_{jd}$, which can in turn affect district outcomes. Appendix Section A.3 shows that similar insights emerge where there are multiple (possibly correlated) compensating differentials.

The asymptotic bias can be avoided if one of the following three conditions holds. First, the

---

[12]Since $Z_{id}-Z_{jd}$ and $\varepsilon_{id}-\varepsilon_{jd}$ are normally distributed, $\mathbb{E}[Z_{idk}-Z_{jdk}|\Delta_d] = \mathbb{E}[Z_{idk}-Z_{jdk}] + \frac{Cov[Z_{idk}-Z_{jdk},\Delta_d]}{\mathbb{V}[\Delta_d]}(\Delta_d-\mathbb{E}[\Delta_d])$. In this application, $\mathbb{E}[Z_{id}-Z_{jd}]=0$, $Cov[Z_{idk}-Z_{jdk},\Delta_d]=\beta\sigma_Z^2$, $\mathbb{V}[\Delta_d] = \beta^2\sigma_Z^2+\sigma_\varepsilon^2$, and $\Delta_d-\mathbb{E}[\Delta_d]=-\alpha$. By independence, $\mathbb{E}[\upsilon_d|\Delta_{id}=0,X_{id}=1,X_{jd}=0]=\mathbb{E}[\upsilon_d|\Delta_{id}=0,X_{id}=0,X_{jd}=1]$.

PCRD estimator consistently recovers τ_{PCRD} when no compensating differential is needed ($\alpha = 0$). This requires voters not to select politicians on the basis of X_{id}. Second, there is no asymptotic bias when the compensating differential does not affect the outcome ($\gamma = 0$). This occurs when $Z_{id} - Z_{jd}$ affects which candidate voters prefer, but Z_{id} does not ultimately affect the district outcome of interest. Third, $\widehat{\tau}_{PCRD}$ consistently estimates τ_{PCRD} when the difference in vote share due to X_{id} is exactly offset by noise ($\alpha + \varepsilon_{1d} - \varepsilon_{0d} = 0$). This knife-edge condition is closely related to the previous condition, since ε_{id} are factors that affect candidate i's vote share without affecting their behavior in office. It is only plausible where the signal to noise ratio $\frac{\sigma_Z^2}{\sigma_\varepsilon^2}$ is sufficiently low that candidate vote shares are predominantly determined by chance, rather than systematic differences in attributes. As I discuss below, theories of voting behavior suggest that none of these conditions usually holds.

3.3.2 How does the direction and magnitude of bias vary?

The asymptotic bias of PCRD estimates can be upward or downward, depending on how X_{id} and Z_{id} affect district outcomes. Where the direction of the effect of each variable agrees—either $\tau, \gamma > 0$ or $\tau, \gamma < 0$—equation (8) shows that $\widehat{\tau}_{PCRD}$ is downwardly biased in magnitude. PCRD designs will thus *underestimate* the effect of X_{id} when both candidate characteristics appeal to voters and both characteristics also lead to better (or worse) district-level outcomes once a politician enters office. This would occur where voters select candidates with characteristics that they correctly anticipate will produce better outcomes in office from the perspective of most voters, such as greater economic performance, security, or redistribution towards a majority group. Intuitively, $\widehat{\tau}_{PCRD}$ captures a lower bound on the effect of X_{id} because differences in Y_d due to electing a candidate possessing desirable characteristic X_{id} are counteracted by electing a candidate possessing relatively less of the also-desirable compensating differential Z_{id}. The partisan affiliation example above illustrates such a case.

PCRD designs instead *overestimate* the effect of X_{id} where the signs of τ and γ disagree. In the gender example above, opposing effects can arise when voters incorrectly believe that women

will perform worse. Opposing effects could similarly occur if vote buying efforts win votes for candidates whose non-programmatic policies later reduce voter welfare.

Each direction of bias creates different challenges for hypothesis testing. Where underestimation occurs, the sign of a PCRD estimate that rejects the null hypothesis remains correct because $\widehat{\tau}_{PCRD}$ is a lower bound. Conversely, a failure to reject the null hypothesis is relatively uninformative because it is consistent with both X_{id} not affecting district outcomes and a positive effect of possessing X_{id} canceling out with a negative effect of possessing relatively less of Z_{id}. These challenges operate in reverse where overestimation occurs.

The second term in equation (8) also illustrates when the asymptotic bias is greatest. The magnitude of the bias increases in γ because the compensating differential has a larger effect on district outcomes, and similarly increases in α because more of the compensating differential is required to ensure close elections. The bias also increases in $\frac{\sigma_Z^2}{\sigma_{\tilde{\varepsilon}}^2}$, as noise becomes relatively less important than candidate characteristics in determining whether an election is close. The influence of β on the size of bias is ambiguous because a greater impact of compensating differentials on candidate vote shares reduces the importance of noise in generating close elections but also reduces the size of the compensating differential needed to overcome the difference in X_{id}.

3.4 Bias in PCRD designs in general

To demonstrate these results more generally, I relax the functional form and distributional assumptions imposed on V_{id} and Y_{id}. First, candidate i's vote share is now an unrestricted function $v(X_{id}, X_{jd}, \mathbf{Z}_{id}, \mathbf{Z}_{jd}, \varepsilon_{id}, \varepsilon_{jd})$, where the realization of two independent and identically distributed shocks $(\varepsilon_{id}, \varepsilon_{jd})$ is independent of all other variables. Second, I assume that potential outcomes are additively separable between X_{id} and $\mathbf{Z}_{id}$, but otherwise allow effects of these variables to vary across districts:

Assumption 4. *Candidate i's potential outcome if elected to office is $Y_{id}(1) = \tau_d X_{id} + g(\mathbf{Z}_{id}) + \upsilon_d$, where υ_d is distributed independently of all other variables.*

Additive separability excludes the possibility that the effect of X_{id} varies with compensating differ-

entials $\mathbf{Z}_{id}$. This assumption is not necessary for asymptotic bias to emerge in PCRD designs, but facilitates a simple decomposition of the bias.

The following proposition establishes the quantity that the PCRD estimator converges to:

Proposition 2. *Under Assumptions 1(a), 2, and 4:*

$$\widehat{\tau}_{PCRD} \xrightarrow{p} \tau_{PCRD} + \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] - \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1].$$

This result demonstrates that PCRD designs generally yield asymptotically biased estimates of the LATE of X_{id}. This bias emerges when $\mathbf{Z}_{id}$ is unconditionally correlated with X_{id} and/or correlations with $\mathbf{Z}_{id}$ are induced or altered by conditioning on close elections when compensating differentials are required for elections to be tied.

To focus on the asymptotic bias introduced by post-treatment conditioning, I impose Assumption 3. Extending the intuitions from the example with a single compensating differential, the following proposition establishes three sufficient conditions for PCRD designs to identify and consistently estimate the LATE of X_{id}:

Proposition 3. *Under Assumptions 1(a)-4, $\widehat{\tau}_{PCRD} \xrightarrow{p} \tau_{PCRD}$ if one of the following conditions holds:*

(i) $V_{id} \perp\!\!\!\perp X_{id}, X_{jd}$ among candidates that could enter close races;

(ii) $\mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] = \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1];$

(iii) whenever $v(1, 0, z, z', \varepsilon_{id}, \varepsilon_{jd}) = v(0, 1, z', z, \varepsilon_{jd}, \varepsilon_{id})$, $z = z'$.

Similarly, the following identification result holds under one of the three preceding conditions:

$$\mathbb{E}[Y_d(1) - Y_d(0)|\Delta_d = 0] = \lim_{v \downarrow 0} \mathbb{E}[Y_{id}|\Delta_{id} = v, X_{id} = 1, X_{jd} = 0] - \lim_{v \downarrow 0} \mathbb{E}[Y_{id}|\Delta_{id} = v, X_{id} = 0, X_{jd} = 1].$$

Condition (i) amounts to ensuring that candidate vote shares are not correlated with X_{id} and X_{jd}; this means that compensating differentials are not required for elections to be close. Condition (ii) allows for compensating differentials to counteract the electoral advantage of X_{id} if they do not

affect district outcomes in expectation. This condition encompasses two cases: no compensating differential affects Y_d; or the net effect of compensating differentials exactly cancels out. Condition (iii) similarly ensures that noise—factors that benefit a candidate which are not features of the candidate themselves—exclusively compensates for the electoral advantage of X_{id} in close races.

4   Implications for applied research

The preceding analysis demonstrated that PCRD designs seeking to isolate effects of an elected politician characteristic require that researchers invoke stronger additional assumptions or accept post-treatment bias (and possibly correlated characteristics bias). This section explores the implications for applied research, starting by considering the viability of imposing one of the conditions in Proposition 3. Since these conditions are unlikely to hold in many applications, I then consider strategies to mitigate the threat to internal validity. I finally discuss the implications of redefining treatment to encompass X_{id} and all compensating differentials.

4.1   Invoking an additional assumption

Perhaps the most appealing method for addressing the inconsistency of PCRD estimates is to explicitly invoke and substantiate one of the conditions in Proposition 3. Beyond imposing Assumption 3 around the discontinuity, this entails assuming—at the discontinuity—that X_{id} does not affect candidate vote shares (condition (i)), that compensating differentials induced by variation in X_{id} do not affect the outcome of interest (condition (ii)), or that compensating differentials are not required because idiosyncratic electoral shocks counterbalance compensating differentials (condition (iii)). I focus on conditions (i) and (ii). Condition (iii) is similar to condition (ii) in claiming that other factors affecting election results do not shape post-election district outcomes; moreover, this knife-edge condition almost surely fails to hold when condition (i) does not hold. Unfortunately, as I next explain, neither condition (i) nor (ii) is easily validated and both assumptions conflict with theories of voting behavior.

4.1.1 Empirical challenges

Where compensating differentials are observed or assumed to exist, showing that no compensating differential affects the outcome of interest is particularly difficult. First, strong support for condition (ii) requires that a researcher further show that *each* compensating differential has no effect on the outcome around the discontinuity. Finding identification strategies for all potential compensating differentials—or even just the most plausible compensating differentials—is unlikely to be feasible. Moreover, if the effects of compensating differentials are heterogeneous with respect to V_{id}, this challenge is exacerbated by the need for these estimates to be local to close elections. Second, because compensating differentials like candidate competence are often difficult to measure, it is hard to confidently claim that unobserved compensating differentials are not affecting the outcome.

Validating that characteristic X_{id} does not affect candidate vote shares around the discontinuity is more attainable. Since X_{id} and V_{id} are both observed, a single test demonstrating that X_{id} does not affect V_{id} among candidates around the discontinuity can support condition (i). A compelling test showing that candidate i's gender or party affiliation does not affect their vote share requires an additional research design exogenously varying X_{id} among candidates that end up in close races. At the expense of external validity, conjoint experiments (Hainmueller, Hopkins and Yamamoto 2014) could help to establish the electoral value of X_{id}. A less compelling test might instead show a limited correlation between X_{id} and V_{id}. However, researchers cannot simply use the *ex post* sample of close elections, where X_{id} and V_{id} are uncorrelated by construction when $\Delta_{id} = 0$. To capture the local effect of X_{id} among candidates that end up in close races, researchers might consider estimating treatment effects among districts where elections were *predicted* to be close.

4.1.2 Theoretical challenges

Compounding the empirical challenge of validating conditions (i) and (ii), both assumptions are often theoretically implausible. Researchers using PCRD designs are usually interested in characteristics like gender or partisan affiliation because they expect these characteristics to impact outcomes that voters also care about. For characteristic X_{id} not to influence candidate vote shares—as

condition (i) requires—when it does affect district outcomes, voters would need to be oblivious to, or not vote on the basis of, the characteristic's expected impact on the outcome of interest to the researcher *or other outcomes voters are concerned about*. For example, Gagliarducci and Paserman (2012) argue that Italian voters are unlikely to vote on the basis of government termination risks and are poorly informed about whether gender might precipitate termination. Nevertheless, termination risk could still correlate with other outcomes that influence vote choice and voters attribute to gender. For condition (ii), the existence of compensating differentials that do not affect the district outcome of interest would require voters to wrongly believe that these characteristics affect an outcome they care about or not care about the outcome of interest to the researcher.

Most theories of voting behavior suggest that at least some voters observe candidate characteristics and understand how such characteristics affect outcomes they care about. Even where electorates are only partially informed about the link between characteristics and outcomes, candidates with identical vote shares should produce identical welfare outcomes—broadly construed to encompass any outcome that matters to voters which different candidates could affect—in expectation when limited information is aggregated across a population (Fowler 2018). Equal vote shares could reflect equally effective candidates committing to policies that converge on the median voter's preferred policy or comparative advantages of one candidate on some dimensions being counteracted by the comparative advantages of other candidates on other dimensions. The latter explanation does not prevent characteristic X_{id} from affecting an outcome of interest, just that other characteristics produce offsetting effects on other outcomes that leave the median voter indifferent between two different candidates. Even if voters are not fully rational, the possibility that easily-observed candidate characteristics that impact important outcomes are simultaneously uncorrelated with candidate vote shares is implausible in many contexts.

4.2 Mitigating threats to internal validity

Where one of conditions (i)-(iii) is not invoked, PCRD designs lack a compelling foundation for identifying the effect of characteristic X_{id}. This is because candidates that narrowly win *must* differ,

in expectation, in other consequential ways too.[13] I next discuss three potential strategies to combat the post-treatment bias inherent to PCRD designs, and their limitations.

4.2.1 Continuity tests

As Table 1 shows, most studies using PCRD designs conduct continuity tests to validate that potentially-confounding district- or candidate-level characteristics do not vary discontinuously at the point of discontinuity. This entails estimating $\lim_{v \downarrow v} \mathbb{E}[Z_{1dk}|\Delta_d = v] - \lim_{v \uparrow v} \mathbb{E}[Z_{0dk}|\Delta_d = v]$ to test the null hypothesis that $\mathbb{E}[Z_{idk}|\Delta_d = 0, X_{id} = 1, X_{jd} = 0] = \mathbb{E}[Z_{idk}|\Delta_d = 0, X_{id} = 0, X_{jd} = 1]$ for each observable covariate k.

In standard RD designs, finding only differences consistent with statistical chance corroborates Assumption 1 (Cattaneo and Titiunik forthcoming; Imbens and Lemieux 2008). In PCRD designs, *district-level* characteristics remain useful for balance tests because—as Sekhon and Titiunik (2012) note—the standard RD continuity assumption implies continuity in district-level characteristics. Intuitively, this is because district-level characteristics can determine the types of races that are close and the degree to which compensating differentials are required by affecting electoral advantages, but cannot vary across candidate types within a given race.[14] District-level continuity tests are rightly common in PCRD applications, but do not imply continuity in candidate-level covariates.

In contrast, continuity tests for *candidate-level* covariates operate differently. If neither condition (i) nor condition (ii) can be invoked,[15] there *must* exist at least one (observable or unobservable) compensating differential. Consequently, detecting discontinuities in theoretically-plausible

---

[13]The asymptotic bias of PCRD designs is smaller than for designs that do not restriction attention to comparisons between relatively similar candidates when the post-treatment bias is small relative to district- and candidate-level differences between districts that differ in X_d; see Appendix section A.4.

[14]For example, assume V_{id} does not depend on $Z_{id} = Z_{jd} = Z_d$ in equation (5). However, a district-level characteristic could *differentially* affect candidate type $X_{id} = 1$, e.g. if $V_{id} = \alpha \frac{X_{id}-X_{jd}}{2} + \beta_1 \frac{(Z_{id1}-Z_{jd1})}{2} + \beta_2 \frac{(X_{id}-X_{jd})Z_{d2}}{2} + \frac{\varepsilon_{id}-\varepsilon_{jd}}{2}$. Nevertheless, the effect of Z_d on Y_d—given by $\mathbb{E}[Z_d|\alpha + \beta_1(Z_{id1} - Z_{jd1}) + \beta_2 Z_{d2} + \varepsilon_{1d} - \varepsilon_{0d} = 0]$—is the same when types $X_{id} = 1$ and $X_{id} = 0$ narrowly win, and thus cancels out. The magnitude of $Z_{id1} - Z_{jd1}$ required to compensate for $\alpha + \beta_2 Z_{d2}$, rather than α, increases though.

[15]Or another condition ensuring $\mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] = \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1]$.

26

compensating differentials—as Gagliarducci and Paserman (2012) do for age, education, and vocational experience—can now serve as a manipulation check guiding researcher interpretation of PCRD estimates. As I discuss below, characterizing compensating differentials—and thus potential sources of bias—can inform efforts to bound estimates or reinterpret treatments. Conversely, failing to reject continuity in observable candidate-level covariates does not necessarily validate a PCRD design. This is because continuity in observable covariates is consistent with condition (i) or (iii) holding *as well as* the existence of unobserved compensating differentials, a lack of statistical power to detect observable compensating differentials, or false positive results.

4.2.2   Covariate adjustment

Where the assumptions necessary for identification do not obviously hold, a strategy common to many statistical approaches is covariate adjustment. This involves adjusting for predetermined potential confounders to the greatest extent possible using observable covariates. Gagliarducci and Paserman (2012) address imbalances between men and women who narrowly won mayoral in elections in Italy by adjusting for various covariates, including those on which significant imbalances were observed. Covariate adjustment can be implemented by adjusting for a subset of compensating differentials $\mathbf{Z}_{id}^{cond} \subset \mathbf{Z}_{id}$ using local polynomial estimators (Calonico et al. 2019).

However, adjusting for candidate-level characteristics does little to address the post-treatment bias that arises in PCRD designs. Since equation (7) must always hold, covariate adjustment does not increase the plausibility of condition (i) because conditioning on $\mathbf{Z}_{id}^{cond}$ induces or accentuates the need for compensating differentials in terms of other covariates $\mathbf{Z}_{id} \setminus \mathbf{Z}_{id}^{cond}$ that are not adjusted for. For example, a researcher using a PCRD design to estimate effects of electing university-educated politicians might condition on ideology because they are concerned that better-educated politicians are in close races because they espouse unpopular policy positions. Even if covariate adjustment breaks the correlation between education and ideology, university-educated politicians in close races with non-university-educated politicians with similar ideologies must still, in expectation, differ in other ways to remain in close races. Covariate adjustment can increase the plausibility

of conditions (ii) and (iii) by increasing the share of variation in candidate vote share explained by noise, i.e. reducing $\frac{\sigma^2_{Z|\mathbf{z}^{cond}}}{\sigma^2_{\tilde{\varepsilon}}}$. However, candidate vote shares only differ due to noise once a researcher has fully adjusted for *all* compensating differentials that affect district-levels outcomes, including compensating differentials induced by conditioning.

4.2.3 Bounding and correcting effect magnitudes

Where none of the additional conditions that yield consistent PCRD estimates are plausible, a more promising strategy—at least for more limited researcher objectives—is to use $\widehat{\tau}_{PCRD}$ to bound the effect of X_{id} or correct estimates of its effects. Such strategies can help establish the direction of the effect or the direction and degree of bias that would nullify or reverse a directional finding.

The preceding discussion of when PCRD designs over and underestimate effects of X_{id} illuminates the benefits and drawbacks to bounding. Indeed, underestimation—which enables researchers to claim that an effect of X_{id} is not smaller in magnitude than $\widehat{\tau}_{PCRD}$—occurs when the net effect of all compensating differentials affects the outcome in the same way as X_{id} at the discontinuity. This relatively strong conclusion for non-null findings could be substantiated by using continuity tests to identify compensating differentials and then providing theoretical or empirical evidence to argue that X_{id} and $\mathbf{Z}_{id}$ affect Y_d in the same direction at the discontinuity. For example, if primary voters are averse to ideological extremists and such candidates must compensate by being more competent on average, then Hall's (2015) results might understate the general election penalty associated with selecting extreme candidates in primary elections. However, establishing the direction of an effect is harder when $\widehat{\tau}_{PCRD}$ fails to reject the null hypothesis because we cannot be sure if underestimation accounts for accepting the null.

In the spirit of Rosenbaum (2002), bias correction may be possible where compensating differentials are observable and plausible estimates of their effects on district outcomes can be imputed. If g is a linear function and $\widehat{\gamma}_k$ is a credible estimate of the LATE of each compensating differential Z_{idk} at the point of discontinuity, then Proposition 2 implies that the PCRD estimate could be

corrected to obtain the LATE of X_{id} as follows:[16]

$$\widehat{\tau}_{PCRD}^{corr} = \widehat{\tau}_{PCRD} - \sum_k \widehat{\gamma}_k \widehat{\delta}_k, \tag{9}$$

where each $\widehat{\delta}_k$ consistently estimates $\mathbb{E}[Z_{idk}|\Delta_d = 0, X_{id} = 1, X_{jd} = 0] - \mathbb{E}[Z_{idk}|\Delta_d = 0, X_{id} = 0, X_{jd} = 1]$ using a (dis)continuity test for observable compensating differentials. Even without estimates of γ_k, researchers could examine the sensitivity of their results to plausible values of γ_k. In the gender example, this could involve estimating the difference in candidate competence between men and women who narrowly win and then $\widehat{\tau}_{PCRD}$ for plausible values of γ_k. Appendix Section A.5 provides examples of how attention to theorized mechanisms can inform γ_k.

4.3 Expanding the conception of treatment

Given the challenges of isolating effects of politician characteristic X_{id}, an alternative approach is to explicitly redefine the estimand to include compensating differentials. Specifically, researchers might target a compound treatment effect incorporating effects of X_{id} *and* all compensating differentials induced or altered by PCRD designs at the point of discontinuity. Hall (2015:24) adopts this type of approach when noting that his PCRD estimate of the effect of selecting ideological extremists in U.S. primary elections on general election outcomes "includes the component of the overall effect that comes from the change in ideology, but also includes any other factors that differ between the two types of candidates." He thus distinguishes between a specific individual characteristic—*extremism*—and the bundle of correlated characteristics characterizing a typical *extremist*. This logic extends to distinguishing the effects of electing a candidate representing the positions of the Democratic party from electing candidates that are Democrats. In the context of women winning primary elections, Bucchianeri (2018:445) similarly defines his estimand as the "causal effect of nominating a female candidate, *not* the causal effect of gender," and notes that this bundle could include compensating differentials that ensure women remain in close races with men. The key

---

[16]Letting $\mathbf{Z}_{id}$ include higher-order polynomials and interactions between characteristics, the Weierstrass approximation theorem ensures that $\sum_k \gamma_k Z_{idk}$ approximates $g(\mathbf{Z}_{id})$.

advantage of redefining the treatment of interest to include correlated characteristics and compensating differentials is that PCRD designs can now yield consistent estimates, albeit for a different estimand.

Formally, this reconceptualization entails focusing on joint potential outcomes $Y_d(X_d, \mathbf{Z}_d) = y(X_d, \mathbf{Z}_d) + \upsilon_d$, where $\mathbf{Z}_d$ remains a vector of other characteristics of the winning candidate in district d. The following proposition characterizes the compound treatment effect that a PCRD estimator converges to under the standard RD continuity assumption:

Proposition 4. *Under Assumptions 1(a) and 2:*

$$\widehat{\tau}_{PCRD} \xrightarrow{p} \int [y(1,z) - y(0,z)] f_c(z) dz + \int y(1,z) f_1(z) dz - \int y(0,z) f_0(z) dz,$$

where $f_{\mathbf{Z}_{id}|\cdot}(z)$ is the conditional probability density function of $\mathbf{Z}_{id}$, the pointwise common component of the density is $f_c(z) := \min \left\{ f_{\mathbf{Z}_{id}|\Delta_{id}=0, X_{id}=1, X_{jd}=0}(z), f_{\mathbf{Z}_{id}|\Delta_{id}=0, X_{id}=0, X_{jd}=1}(z) \right\}$, and $f_m(z) := f_{\mathbf{Z}_{id}|\Delta_{id}=0, X_{id}=m, X_{jd}=1-m}(z) - f_c(z)$ is the excess density among politicians of type $X_{id} = m$ that win close elections.

When the distribution of $\mathbf{Z}_{id}$ differs across candidates of type $X_{id} = 1$ and $X_{id} = 0$ that win close elections, this result shows that $\widehat{\tau}_{PCRD}$ captures effects of both X_{id} and $\mathbf{Z}_{id}$. The proposition expresses this in terms of a LATE of X_{id}, weighted by the common distribution of $\mathbf{Z}_{id}$ characteristics across candidate types, *and* effects of differences in the distribution of $\mathbf{Z}_{id}$ across candidate that differ in X_{id}. For Hall (2015), X_{id} represents ideological extremism and $\mathbf{Z}_{id}$ captures all other characteristics of extremists—both those that naturally correlate with X_{id} *and* those induced, accentuated, or attenuated by conditioning the estimand on close elections that are affected by X_{id}.

Reconceptualizing potential confounders as part of the PCRD estimand again implies a nonstandard role for candidate-level continuity tests. Rather than validating Assumption 1, candidate-level covariate tests yield estimates of $\widehat{\delta}_k$ that now help to characterize the compound treatment. Substantial differences in $Z_{1dk} - Z_{0dk}$ at the point of discontinuity suggest that Z_{idk} may be an important component of the comparison captured by $\widehat{\tau}_{PCRD}$, whereas the reverse holds for covariates

where discontinuities are not detected. Hall (2015) adopts such an approach by examining whether extremist primary winners differ from non-extremist winners on other dimensions in his analysis of mechanisms.

There are, however, three notable drawbacks to broadening the notion of treatment; the importance of each drawback will vary by application. First, it is hard to fully characterize treatment in many empirical applications. This is both because compensating differentials should generally exist but continuity tests may lack the statistical power to detect differences in $\mathbf{Z}_{id}$ and because researchers may struggle to adequately measure relevant elements of $\mathbf{Z}_{id}$. Beyond the label "X and all its compensating differentials," PCRD designs lack clarity about the bundle of characteristics that constitute the treatment.

Second, the external validity and interpretability of PCRD estimates may be limited because the design is unlikely to capture typical or homogeneous bundles of characteristics. PCRD designs identify compound treatments defined by the correlations between characteristics that exist *after* conditioning on close elections where X_{id} affects vote shares. Extremists that win close races against non-extremists may thus be atypical of extremists that narrowly win any type of primary election, in addition to being atypical of extremists in general. Moreover, because many permutations of $(\mathbf{Z}_{id}, \mathbf{Z}_{jd})$ can produce close elections, narrowly winning candidates of type X_{id} can experience different values of $\mathbf{Z}_{id}$—a violation of the treatment uniformity component of SUTVA. For example, some extremists that overcome an electoral penalty associated with their ideological extremism to win may be unusually competent and others may offer more appealing platforms.

Third, bundled treatments limit the degree to which PCRD designs can test specific theories or inform certain policy decisions. Theories often specify "all else equal" comparative statics for different variables that PCRD designs cannot distinguish because all candidate-level characteristics, albeit to differing degrees, are considered part of a compound treatment. PCRD designs therefore cannot reveal whether ideological extremists lose general elections because of their policy positions, differences in competence between extremist and non-extremist candidates that narrowly win, or some combination of both.

The extent to which this inability to distinguish the contribution of different elements of the compound treatment limits the relevance of PCRD estimates to policymakers likely depends on the policy tools available. On one hand, policymakers with constrained choice sets may not care which part of the treatment matters, only that policies that encourage (or discourage) politicians of type X_{id} to run for office or help such candidates win elections should be favored. For example, local party committees might alter candidacy rules to avoid selecting extremist candidates that lose general elections. On the other hand, the limited information conveyed by PCRD estimates is less helpful where policymakers are picking between or devising more fine-grained policies that can encourage candidates of type X_{id} instead of type Z_{idk}. Understanding the mechanism driving PCRD estimates could be consequential for reformers investigating whether they should adopt gender quotas or require more specific competencies of their candidates.

5   Conclusions

This article demonstrates that PCRD designs—a popular approach used to estimate effects of a specific characteristic, or bundle of characteristics, of elected politicians on downstream outcomes—generally require imposing substantially stronger assumptions than standard RD designs. This is because the treatment variable in this non-standard RD application is defined both by winning close elections and a candidate characteristic that can affect selection into the set of narrow election winners of different types. I have shown that such post-treatment conditioning causes PCRD designs to capture the effect of the specific characteristic of interest together with all the compensating differentials required for candidates with the characteristic of interest to remain in close races.

Even when the characteristic of interesting is unconditionally independent of other characteristics, PCRD designs generate inconsistent estimates of the LATE exclusively attributable to the characteristic that defines treatment, except under two strong additional assumptions: (i) the characteristic of interest did not affect the candidate's vote share; or (ii) no compensating differential affected the outcome of interest. Unfortunately, neither condition is plausible in many contexts

and both are difficult to empirically validate. Accordingly, PCRD designs cannot generally isolate impacts of specific politician characteristics on outcomes relating to political representation, accountability, and participation.

Researchers can attempt to combat this challenge in several ways. One approach is to explicitly accept and then mitigate threats to internal validity by using theory and data to bound or sign treatment effects. Another approach broadens the definition of treatment to exclude the possibility of candidate-level confounding by redefining the estimand to include both the characteristic of interest and all compensating differentials. Both approaches entail trade-offs, either in terms of internal validity or the generality of treatment, but could nevertheless illuminate hypotheses in certain settings—even though PCRD designs cannot isolate effects in the same way as standard RD designs.

References

Abou-Chadi, Tarik and Werner Krause. 2020. "The causal effect of radical right success on mainstream parties' policy positions: A regression discontinuity approach." *British Journal of Political Science* 50(3):829–847.

Akhtari, Mitra, Diana Moreira and Laura Trucco. 2022. "Political turnover, bureaucratic turnover, and the quality of public services." *American Economic Review* 112(2):442–493.

Aksoy, Ozan and Diego Gambetta. 2021. "The politics behind the veil." *European Sociological Review* 37(1):67–88.

Aksoy, Ozan and Francesco C. Billari. 2018. "Political Islam, marriage, and fertility: evidence from a natural experiment." *American Journal of Sociology* 123(5):1296–1340.

Alberti, Carla, Diego Diaz-Rioseco and Giancarlo Visconti. forthcoming. "Gendered bureaucracies: Women mayors and the size and composition of local governments." *Governance* .

Albouy, David. 2011. "Do voters affect or elect policies? A new perspective, with evidence from the US Senate." *Electoral Studies* 30(1):162–173.

Albouy, David. 2013. "Partisan representation in Congress and the geographic distribution of federal funds." *Review of Economics and Statistics* 95(1):127–141.

Alonso, José M and Rhys Andrews. 2020. "Political ideology and social services contracting: evidence from a regression discontinuity design." *Public Administration Review* 80(5):743–754.

Anastasopoulos, Lefteris. 2016. "Estimating the gender penalty in House of Representative elections using a regression discontinuity design." *Electoral Studies* 43:150–157.

Anzia, Sarah F. and Christopher R. Berry. 2011. "The Jackie (and Jill) Robinson effect: why do congresswomen outperform congressmen?" *American Journal of Political Science* 55(3):478–493.

Aragón, Fernando and Ricardo Pique. 2020. "Better the devil you know? Reelected politicians and policy outcomes under no term limits." *Public Choice* 182(1):1–16.

Arvate, Paulo and Joana Story. forthcoming. "Leaders affect business creation: Evidence from mayoral elections." *Leadership Quarterly* .

Arvate, Paulo, Sergio Firpo and Renan Pieri. 2017. "Future electoral impacts of having a female mayor." *Brazilian Political Science Review* 11(2):1.

Arvate, Paulo, Sergio Firpo and Renan Pieri. 2021. "Can women's performance in elections determine the engagement of adolescent girls in politics?" *European Journal of Political Economy* 70:102045.

Asher, Sam and Paul Novosad. 2017. "Politics and local economic growth: Evidence from India." *American Economic Journal: Applied Economics* 9(1):229–73.

Ashworth, Scott, Christopher R. Berry and Ethan Bueno de Mesquita. 2021. "Modeling Theories of Women's Underrepresentation in Elections." Working paper.

Barber, Michael, Daniel M. Butler and Jessica Preece. 2016. "Gender inequalities in campaign finance." *Quarterly Journal of Political Science* 11(2):219–48.

Baskaran, Thushyanthan and Zohal Hessami. 2018. "Does the election of a female leader clear the way for more women in politics?" *American Economic Journal: Economic Policy* 10(3):95–121.

Bauhr, Monika and Nicholas Charron. 2021. "Will women executives reduce corruption? Marginalization and network inclusion." *Comparative Political Studies* 54(7):1292–1322.

Beach, Brian and Daniel B. Jones. 2017. "Gridlock: Ethnic diversity in government and the provision of public goods." *American Economic Journal: Economic Policy* 9(1):112–36.

Beland, Louis-Philippe. 2015. "Political parties and labor-market outcomes: Evidence from US states." *American Economic Journal: Applied Economics* 7(4):198–220.

Beland, Louis-Philippe and Bulent Unel. 2018. "The impact of party affiliation of US governors on immigrants' labor market outcomes." *Journal of Population Economics* 31(2):627–670.

Beland, Louis-Philippe and Bulent Unel. 2019. "Politics and Entrepreneurship in the US." *Canadian Journal of Economics/Revue canadienne d'économique* 52(1):33–57.

Beland, Louis-Philippe and Sara Oloomi. 2017. "Party affiliation and public spending: Evidence from US governors." *Economic Inquiry* 55(2):982–995.

Beland, Louis-Philippe and Vincent Boucher. 2015. "Polluting politics." *Economics Letters* 137:176–181.

Bhalotra, Sonia and Irma Clots-Figueras. 2014. "Health and the political agency of women." *American Economic Journal: Economic Policy* 6(2):164–97.

Bhalotra, Sonia, Irma Clots-Figueras, Guilhem Cassan and Lakshmi Iyer. 2014. "Religion, politician identity and development outcomes: Evidence from India." *Journal of Economic Behavior & Organization* 104:4–17.

Bhalotra, Sonia, Irma Clots-Figueras and Lakshmi Iyer. 2018. "Pathbreakers? Women's electoral success and future political participation." *Economic Journal* 128(613):1844–1878.

Bhalotra, Sonia, Irma Clots-Figueras and Lakshmi Iyer. 2021. "Religion and abortion: The role of politician identity." *Journal of Development Economics* 153:102746.

Bracco, Emanuele, Ben Lockwood, Francesco Porcelli and Michela Redoano. 2015. "Intergovernmental grants as signals and the alignment effect: Theory and evidence." *Journal of Public Economics* 123:78–91.

Bracco, Emanuele, Francesco Porcelli and Michela Redoano. 2019. "Political competition, tax salience and accountability: Theory and evidence from Italy." *European Journal of Political Economy* 58:138–163.

Bracco, Emanuele, Maria De Paola, Colin P. Green and Vincenzo Scoppa. 2018. "The effect of far right parties on the location choice of immigrants: Evidence from Lega Nord Mayors." *Journal of Public Economics* 166:12–26.

Brollo, Fernanda, Katja Kaufmann and Eliana La Ferrara. 2020. "The political economy of program enforcement: Evidence from Brazil." *Journal of the European Economic Association* 18(2):750–791.

Brollo, Fernanda and Tommaso Nannicini. 2012. "Tying your enemy's hands in close races: the politics of federal transfers in Brazil." *American Political Science Review* 106(4):742–761.

Brollo, Fernanda and Ugo Troiano. 2016. "What happens when a woman wins an election? Evidence from close races in Brazil." *Journal of Development Economics* 122:28–45.

Broockman, David E. 2014. "Do female politicians empower women to vote or run for office? A regression discontinuity approach." *Electoral Studies* 34:190–204.

Bruce, Raphael, Alexsandros Cavgias, Luis Meloni and Mário Remígio. 2022. "Under pressure: Women's leadership during the COVID-19 crisis." *Journal of Development Economics* 154:102761.

Bucchianeri, Peter. 2018. "Is Running Enough? Reconsidering the Conventional Wisdom about Women Candidates." *Political Behavior* 40(2):435–466.

Bueno de Mesquita, Ethan and Scott A. Tyson. 2020. "The commensurability problem: Conceptual difficulties in estimating the effect of behavior on behavior." *American Political Science Review* 114(2):375–391.

Bueno, Natália S. 2018. "Bypassing the enemy: Distributive politics, credit claiming, and nonstate organizations in brazil." *Comparative Political Studies* 51(3):304–340.

Bullock, John G., Donald P. Green and Shang E. Ha. 2010. "Yes, but what's the mechanism?(don't expect an easy answer)." *Journal of personality and social psychology* 98(4):550.

Butler, Daniel M. 2009. "A regression discontinuity design analysis of the incumbency advantage and tenure in the US House." *Electoral Studies* 28(1):123–128.

Callen, Michael, Saad Gulzar and Arman Rezaee. 2020. "Can political alignment be costly?" *Journal of Politics* 82(2):612–626.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell and Rocío Titiunik. 2019. "Regression discontinuity designs using covariates." *Review of Economics and Statistics* 101(3):442–451.

Calonico, Sebastian, Matias D. Cattaneo and Rocío Titiunik. 2014. "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica* 82(6):2295–2326.

Carnes, Nicholas and Noam Lupu. 2016. "What good is a college degree? Education and leader quality reconsidered." *Journal of Politics* 78(1):35–49.

Carter, Christopher L. 2020. "Party system erosion: Evidence from Peru." *Party Politics* 26(5):581–593.

Casarico, Alessandra, Salvatore Lattanzio and Paola Profeta. forthcoming. "Women and local public finance." *European Journal of Political Economy* .

Cattaneo, Matias D. and Rocío Titiunik. forthcoming. "Regression Discontinuity Designs." *Annual Review of Economics* .

Caughey, Devin, Christopher Warshaw and Yiqing Xu. 2017. "Incremental democracy: The policy effects of partisan control of state government." *Journal of Politics* 79(4):1342–1358.

Caughey, Devin and Jasjeet S. Sekhon. 2011. "Elections and the regression discontinuity design: Lessons from close US house races, 1942-2008." *Political Analysis* 19(4):385–408.

Chemin, Matthieu. 2012. "Welfare Effects of Criminal Politicians: A Discontinuity-Based Approach." *Journal of Law and Economics* 55(3):667–690.

Cheng, Chao-Yo and Johannes Urpelainen. 2019. "Criminal Politicians and Socioeconomic Development: Evidence from Rural India." *Studies in Comparative International Development* 54:501–527.

Christensen, Darin and Francisco Garfias. 2021. "The politics of property taxation: Fiscal infrastructure and electoral incentives in Brazil." *Journal of Politics* 83(4):1399–1416.

Clots-Figueras, Irma. 2011. "Women in politics: Evidence from the Indian States." *Journal of Public Economics* 95(7-8):664–690.

Clots-Figueras, Irma. 2012. "Are female leaders good for education? Evidence from India." *American Economic Journal: Applied Economics* 4(1):212–244.

Corekcioglu, Gozde. 2021. "Unveiling the effects of a headscarf ban: Evidence from municipal jobs in Turkey." *Journal of Comparative Economics* 49(2):382–404.

Coviello, Decio and Stefano Gagliarducci. 2017. "Tenure in office and public procurement." *American Economic Journal: Economic Policy* 9(3):59–105.

Cunial, Santiago L. 2021. "Transitions for whom? Political alignment and subsidies for solar energy projects in rural Colombian municipalities." *Latin American Policy* 12(2):300–332.

Curto-Grau, Marta, Albert Solé-Ollé and Pilar Sorribas-Navarro. 2018. "Does electoral competition curb party favoritism?" *American Economic Journal: Applied Economics* 10(4):378–407.

Daly, Sarah Zukerman. forthcoming. "How Do Violent Politicians Govern? The Case of Paramilitary-Tied Mayors in Colombia." *British Journal of Political Science* .

Dasgupta, Aditya and Devesh Kapur. 2020. "The political economy of bureaucratic overload: Evidence from rural development officials in India." *American Political Science Review* 114(4):1316–1334.

de Benedictis-Kessner, Justin and Christopher Warshaw. 2016. "Mayoral partisanship and municipal fiscal policy." *Journal of Politics* 78(4):1124–1138.

de Benedictis-Kessner, Justin and Christopher Warshaw. 2020. "Politics in forgotten governments: the partisan composition of county legislatures and county fiscal policies." *Journal of Politics* 82(2):460–475.

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* 19:375–396.

Dell, Melissa. 2015. "Trafficking networks and the Mexican drug war." *American Economic Review* 105(6):1738–79.

Dynes, Adam M and John B. Holbein. 2020. "Noisy retrospection: The effect of party control on policy outcomes." *American Political Science Review* 114(1):237–257.

Eggers, Andrew C. 2017. "Quality-based explanations of incumbency effects." *Journal of Politics* 79(4):1315–1328.

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

Eggers, Andrew C. and Jens Hainmueller. 2009. "MPs for sale? Returns to office in postwar British politics." *American Political Science Review* 103(4):513–533.

Eggers, Andrew C., Ronny Freier, Veronica Grembi and Tommaso Nannicini. 2018. "Regression discontinuity designs based on population thresholds: Pitfalls and solutions." *American Journal of Political Science* 62(1):210–229.

Erikson, Robert S., Olle Folke and James M. Snyder, Jr. 2015. "A gubernatorial helping hand? How governors affect presidential elections." *Journal of Politics* 77(2):491–504.

Estache, Antonio, Grégoire Garsous and Ronaldo Seroa da Motta. 2016. "Shared mandates, moral

hazard, and political (mis) alignment in a decentralized economy." *World Development* 83:98–110.

Fergusson, Leopoldo, Pablo Querubin, Nelson A. Ruiz and Juan F. Vargas. 2021. "The Real Winner's Curse." *American Journal of Political Science* 65(1):52–68.

Ferraz, Claudio and Frederico Finan. 2011. "Electoral accountability and corruption: Evidence from the audits of local governments." *American Economic Review* 101(4):1274–1311.

Ferreira, Fernando and Joseph Gyourko. 2009. "Do political parties matter? Evidence from US cities." *Quarterly Journal of Economics* 124(1):399–422.

Ferreira, Fernando and Joseph Gyourko. 2014. "Does gender matter for political leadership? The case of US mayors." *Journal of Public Economics* 112:24–39.

Fiva, Jon H and Askill H. Halse. 2016. "Local favoritism in at-large proportional representation systems." *Journal of Public Economics* 143:15–26.

Fiva, Jon H., Olle Folke and Rune J. Sørensen. 2018. "The power of parties: evidence from close municipal elections in Norway." *Scandinavian Journal of Economics* 120(1):3–30.

Folke, Olle. 2014. "Shades of brown and green: party effects in proportional election systems." *Journal of the European Economic Association* 12(5):1361–1395.

Fowler, Anthony. 2018. "A Bayesian explanation for the effect of incumbency." *Electoral Studies* 53:66–78.

Fowler, Anthony and Andrew B. Hall. 2015. "Congressional seniority and pork: A pig fat myth?" *European Journal of Political Economy* 40:42–56.

Fredriksson, Per G., Le Wang and Khawaja A. Mamun. 2011. "Are politicians office or policy motivated? The case of US governors' environmental policies." *Journal of Environmental Economics and Management* 62(2):241–253.

Fredriksson, Per G., Le Wang and Patrick L. Warren. 2013. "Party politics, governors, and economic policy." *Southern Economic Journal* 80(1):106–126.

Freier, Ronny and Christian Odendahl. 2015. "Do parties matter? Estimating the effect of political power in multi-party systems." *European Economic Review* 80:310–328.

Freier, Ronny and Sebastian Thomasius. 2016. "Voters prefer more qualified mayors, but does it matter for public finances? Evidence for Germany." *International Tax and Public Finance* 23(5):875–910.

Gagliarducci, Stefano and M. Daniele Paserman. 2012. "Gender interactions within hierarchies: evidence from the political arena." *Review of Economic Studies* 79(3):1021–1052.

Galindo-Silva, Hector. 2015. "New parties and policy outcomes: Evidence from Colombian local governments." *Journal of Public Economics* 126:86–103.

Gelman, Andrew and Guido Imbens. 2019. "Why high-order polynomials should not be used in regression discontinuity designs." *Journal of Business & Economic Statistics* 37(3):447–456.

Gerber, Elisabeth R. and Daniel J. Hopkins. 2011. "When mayors matter: estimating the impact of mayoral partisanship on city policy." *American Journal of Political Science* 55(2):326–339.

Girardi, Daniele. 2020. "Partisan shocks and financial markets: Evidence from close national elections." *American Economic Journal: Applied Economics* 12(4):224–52.

Gunderson, Anna. forthcoming. "Who punishes more? partisanship, punitive policies, and the puzzle of democratic governors." *Political Research Quarterly* .

Hahn, Jinyong, Petra Todd and Wilbert Van der Klaauw. 2001. "Identification and estimation of treatment effects with a regression-discontinuity design." *Econometrica* 69(1):201–209.

Hainmueller, Jens, Daniel J. Hopkins and Teppei Yamamoto. 2014. "Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments." *Political Analysis* 22(1):1–30.

Hall, Andrew B. 2015. "What happens when extremists win primaries?" *American Political Science Review* 109(1):18–42.

Hall, Andrew B. and Daniel M. Thompson. 2018. "Who Punishes Extremist Nominees? Candidate Ideology and Turning Out the Base in US Elections." *American Political Science Review* 112(3):509–524.

Hankins, William B. and Gary A. Hoover. 2019. "Do Democratic governors lower economic freedom? A regression discontinuity approach." *Journal of Public Finance and Public Choice* 34(2):101–126.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1):153–161.

Heide-Jørgensen, Tobias. 2021. "Triggering Ideological Thinking: How Elections Foster Coherence of Welfare State Attitudes." *American Political Science Review* 115(2):506–521.

Hernán, Miguel A. and James M. Robins. 2011. *Causal inference*. CRC Press.

Hernández Company, José Antonio and David O. Argente Amaya. 2018. "Contextual Incentives are Not Enough: Clientelistic Capacity and the Politics of Enrollment in Mexico's Seguro Popular." *Latin American Policy* 9(2):304–330.

Hopkins, Daniel J. and Katherine T. McCabe. 2012. "After it's too late: estimating the policy impacts of black mayoralties in US Cities." *American Politics Research* 40(4):665–700.

Hyytinen, Ari, Jaakko Meriläinen, Tuukka Saarimaa, Otto Toivanen and Janne Tukiainen. 2018. "Public employees as politicians: Evidence from close elections." *American Political Science Review* 112(1):68–81.

Imbens, Guido W. and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics* 2(142):615–635.

Innes, Robert and Arnab Mitra. 2015. "Parties, politics, and regulation: Evidence from clean air act enforcement." *Economic Inquiry* 53(1):522–539.

Jankowski, Michael, Kamil Marcinkiewicz and Anna Gwiazda. 2019. "The Effect of Electing Women on Future Female Candidate Selection Patterns: Findings from a Regression Discontinuity Design." *Politics & Gender* pp. 1–29.

Johannessen, Peter G. 2020. "Linkage switches in local elections: evidence from the Workers' Party in Brazil." *Comparative Political Studies* 53(1):109–143.

Joshi, Nayan Krishna. 2015. "Party politics, governors, and healthcare expenditures." *Economics & Politics* 27(1):53–77.

Kim, Sung Eun and Johannes Urpelainen. 2017. "The polarization of American environmental policy: A regression discontinuity analysis of Senate and House votes, 1971-2013." *Review of Policy Research* 34(4):456–484.

Kirkland, Patricia A. 2020. "Mayoral Candidates, Social Class, and Representation in American Cities." *Journal of Political Institutions and Political Economy* 1(1):105–136.

Kirkland, Patricia A. 2021. "Business Owners and Executives as Politicians: The Effect on Public Policy." *Journal of Politics* 83(4):1652–1668.

Kogan, Vladimir, Stéphane Lavertu and Zachary Peskowitz. 2021. "How does minority political representation affect school district administration and student outcomes?" *American Journal of Political Science* 65(3):699–716.

Kresch, Evan Plous and Rodrigo Schneider. 2020. "Political determinants of investment in water and sanitation: Evidence from Brazilian elections." *Economics Letters* 189:109041.

Kuipers, Nicholas, Gareth Nellis and Michael Weaver. forthcoming. "Does Electing Islamists Increase Religious Violence and Intolerance?" *British Journal of Political Science* .

Kuk, John and Zoltan Hajnal. 2021. "Democratic Party Control Reduces Gender Inequality." *Legislative Studies Quarterly* 46(1):155–188.

Kuliomina, Jekaterina. 2021. "Do personal characteristics of councilors affect municipal budget allocation?" *European Journal of Political Economy* 70:102034.

Lara E., Bernardo and Sergio Toro M. 2019. "Tactical distribution in local funding: The value of an aligned mayor." *European Journal of Political Economy* 56:74–89.

Lawless, Jennifer L. 2015. "Female Candidates and Legislators." *Annual Review of Political Science* 18:349–366.

Lee, David S., Enrico Moretti and Matthew J. Butler. 2004. "Do voters affect or elect policies? Evidence from the US House." *Quarterly Journal of Economics* 119(3):807–859.

Leigh, Andrew. 2008. "Estimating the impact of gubernatorial partisanship on policy settings and economic outcomes: A regression discontinuity approach." *European Journal of Political Economy* 24(1):256–268.

Lewis, Blane D., Hieu T.M. Nguyen and Adrianus Hendrawan. 2020. "Political accountability and public service delivery in decentralized Indonesia: Incumbency advantage and the performance of second term mayors." *European Journal of Political Economy* 64:101910.

Macartney, Hugh and John D. Singleton. 2018. "School boards and student segregation." *Journal of Public Economics* 164:165–182.

Malik, Rabia. 2021. "(A) political constituency development funds: Evidence from Pakistan." *British Journal of Political Science* 51(3):963–980.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142(2):698–714.

Meyer, Andrew G. 2019. "Elite influence on climate change skepticism: evidence from close gubernatorial elections." *Journal of the Association of Environmental and Resource Economists* 6(4):783–822.

Meyersson, Erik. 2014. "Islamic Rule and the Empowerment of the Poor and Pious." *Econometrica* 82(1):229–269.

Migueis, Marco. 2013. "The effect of political alignment on transfers to Portuguese municipalities." *Economics & Politics* 25(1):110–133.

Natividad, Gabriel. forthcoming. "Decentralizing investment: Evidence from municipal organization after close elections." *Journal of Economics & Management Strategy* .

Nellis, Gareth, Michael Weaver and Steven Rosenzweig. 2016. "Do parties matter for ethnic violence? Evidence from India." *Quarterly Journal of Political Science* 11(3):249–277.

Nellis, Gareth and Niloufer Siddiqui. 2018. "Secular Party Rule and Religious Violence in Pakistan." *American Political Science Review* 112(1):49–67.

Pacca, Lucia, Daniele Curzi, Gordon Rausser and Alessandro Olper. 2021. "The Role of Party Affiliation, Lobbying, and Electoral Incentives in Decentralized US State Support of the Environment." *Journal of the Association of Environmental and Resource Economists* 8(3):617–653.

Payson, Julia A. 2020. "The Partisan Logic of City Mobilization: Evidence from State Lobbying Disclosures." *American Political Science Review* 114(3):677–690.

Peskowitz, Zachary and Suhas A. Sridharan. 2018. "Partisanship, political institutions, and debt issues." *Journal of Law, Economics, and Organization* 34(3):395–424.

Pettersson-Lidbom, Per. 2008. "Do parties matter for economic outcomes? A regression-discontinuity approach." *Journal of the European Economic Association* 6(5):1037–1056.

Potrafke, Niklas and Felix Roesel. 2020. "The urban–rural gap in healthcare infrastructure: does government ideology matter?" *Regional Studies* 54(3):340–351.

Prakash, Nishith, Marc Rockmore and Yogesh Uppal. 2019. "Do criminally accused politicians affect economic outcomes? Evidence from India." *Journal of Development Economics* 141.

Priyanka, Sadia. 2020. "Do female politicians matter for female labor market outcomes? Evidence from state legislative elections in India." *Labour Economics* 64:101822.

Ravanilla, Nico, Renard Sexton and Dotan Haim. forthcoming. "Deadly Populism: How Local Political Outsiders Drive Duterte's War on Drugs in the Philippines." *Journal of Politics* .

Rocha, Fabiana, Veronica Ines Fernandez Orellano and Karina Bugarin. 2018. "Elected in a close race: Mayor's characteristics and local public finances." *EconomiA* 19(2):149–163.

Rosenbaum, Paul R. 2002. *Observational Studies*. Springer.

Sekhon, Jasjeet S. and Rocío Titiunik. 2012. "When natural experiments are neither natural nor experiments." *American Political Science Review* 106(1):35–57.

Solé-Ollé, Albert and Elisabet Viladecans-Marsal. 2013. "Do political parties matter for local land use policies?" *Journal of Urban Economics* 78:42–56.

Szakonyi, David. 2021. "Private Sector Policy Making: Business Background and Politicians' Behavior in Office." *Journal of Politics* 83(1):260–276.

Thompson, Daniel M. 2020. "How partisan is local law enforcement? evidence from sheriff cooperation with immigration authorities." *American Political Science Review* 114(1):222–236.

Vogl, Tom S. 2014. "Race and the politics of close elections." *Journal of Public Economics* 109:101–113.

Walter, André. 2021. "Socialist Threat? Radical Party Entry, Electoral Alliances, and the Introduction of Proportional Representation." *American Political Science Review* 115(2):701–708.

Wang, Kunyu and Anthony Heyes. forthcoming. "Does the party in power affect FDI? First causal evidence from narrow margin US state elections." *Party Politics* .

Wong, Stan Hok-Wui. 2014. "Resource disparity and multi-level elections in competitive authoritarian regimes: Regression discontinuity evidence from Hong Kong." *Electoral Studies* 33:200–219.

Xu, Yiqing and Yang Yao. 2015. "Informal institutions, collective action, and public investment in rural China." *American Political Science Review* 109(2):371–391.

Appendix

Contents

A.1 Published articles using PCRD designs

Table A1 lists 126 published articles using PCRD designs.

Table A1: Studies using PCRD designs to estimate effects of winning candidate characteristics

Candidate characteristic	List of published articles
Criminal history	Chemin (2012), Cheng and Urpelainen (2019), Daly (forthcoming), Prakash, Rockmore and Uppal (2019).
Education	Carnes and Lupu (2016), Freier and Thomasius (2016), Kuliomina (2021), Rocha, Orellano and Bugarin (2018).
Gender	Alberti, Diaz-Rioseco and Visconti (forthcoming), Anastasopoulos (2016), Anzia and Berry (2011), Arvate, Firpo and Pieri (2017), Arvate, Firpo and Pieri (2021), Barber, Butler and Preece (2016), Baskaran and Hessami (2018), Bauhr and Charron (2021), Bhalotra and Clots-Figueras (2014), Bhalotra, Clots-Figueras and Iyer (2018, 2021), Brollo and Troiano (2016), Broockman (2014), Bruce et al. (2022), Bucchianeri (2018), Casarico, Lattanzio and Profeta (forthcoming), Clots-Figueras (2011, 2012), Ferreira and Gyourko (2014), Gagliarducci and Paserman (2012), Jankowski, Marcinkiewicz and Gwiazda (2019), Kuliomina (2021), Priyanka (2020), Rocha, Orellano and Bugarin (2018).
Ideology	Hall (2015), Hall and Thompson (2018).
Incumbency, term limit status, or seniority	Akhtari, Moreira and Trucco (2022), Aragón and Pique (2020),Brollo, Kaufmann and La Ferrara (2020), Butler (2009), Christensen and Garfias (2021), Coviello and Gagliarducci (2017), Ferraz and Finan (2011), Fowler and Hall (2015), Freier and Thomasius (2016), Lewis, Nguyen and Hendrawan (2020), Rocha, Orellano and Bugarin (2018).
Partisan alignment across levels of government	Asher and Novosad (2017), Bracco et al. (2015), Brollo and Nannicini (2012), Bueno (2018), Callen, Gulzar and Rezaee (2020), Cunial (2021), Curto-Grau, Solé-Ollé and Sorribas-Navarro (2018), Dasgupta and Kapur (2020), Dell (2015), Erikson, Folke and Snyder (2015), Estache, Garsous and da Motta (2016), Kresch and Schneider (2020), Lara E. and Toro M. (2019), Malik (2021), Migueis (2013), Payson (2020), Ravanilla, Sexton and Haim (forthcoming).
Partisan affiliation	Abou-Chadi and Krause (2020), Aksoy and Billari (2018), Aksoy and Gambetta (2021), Albouy (2011, 2013), Alonso and Andrews (2020), Arvate and Story (forthcoming), Beland (2015), Beland and Boucher (2015), Beland and Oloomi (2017), Beland and Unel (2018, 2019), Bracco et al. (2018); Bracco, Porcelli and Redoano (2019), Carter (2020), Caughey, Warshaw and Xu (2017), Corekcioglu (2021), de Benedictis-Kessner and Warshaw (2016, 2020), Dynes and Holbein (2020), Fergusson et al. (2021), Ferreira and Gyourko (2009), Fiva, Folke and Sørensen (2018), Fiva and Halse (2016), Folke (2014), Fredriksson, Wang and Mamun (2011), Fredriksson, Wang and Warren (2013), Freier and Odendahl (2015), Galindo-Silva (2015), Gerber and Hopkins (2011), Girardi (2020), Gunderson (forthcoming), Hankins and Hoover (2019), Heide-Jørgensen (2021), Hernández Company and Argente Amaya (2018), Innes and Mitra (2015), Johannessen (2020), Joshi (2015), Kim and Urpelainen (2017), Kuipers, Nellis and Weaver (forthcoming), Kuk and Hajnal (2021), Lee, Moretti and Butler (2004), Leigh (2008), Macartney and Singleton (2018), Meyer (2019), Meyersson (2014), Natividad (forthcoming), Nellis and Siddiqui (2018), Nellis, Weaver and Rosenzweig (2016), Pacca et al. (2021), Peskowitz and Sridharan (2018), Pettersson-Lidbom (2008), Potrafke and Roesel (2020), Solé-Ollé and Viladecans-Marsal (2013), Thompson (2020), Walter (2021), Wang and Heyes (forthcoming), Wong (2014).
Pre-office vocation	Hyytinen et al. (2018), Kirkland (2020, 2021), Kuliomina (2021), Szakonyi (2021).
Race, ethnicity, religion, or clan	Beach and Jones (2017), Bhalotra et al. (2014), Bhalotra, Clots-Figueras and Iyer (2021), Hopkins and McCabe (2012), Kogan, Lavertu and Peskowitz (2021), Vogl (2014), Xu and Yao (2015).

Note: This list excludes unpublished studies and may be incomplete.

A.2 Proof of propositions

Proof of Proposition 1: The following derivation establishes consistency:

$$
\begin{aligned}
\widehat{\tau}_{RD} &= \widehat{\mu}_+(0) - \widehat{\mu}_-(0) \\
&\xrightarrow{p} \lim_{v \downarrow 0} \mathbb{E}[Y_{id}|\Delta_{id} = v] - \lim_{v \uparrow 0} \mathbb{E}[Y_{id}|\Delta_{id} = v] \\
&= \lim_{v \downarrow 0} \mathbb{E}[Y_{id}(1)|\Delta_{id} = v] - \lim_{v \uparrow 0} \mathbb{E}[Y_{id}(0)|\Delta_{id} = v] \\
&= \mathbb{E}[Y_{id}(1)|\Delta_{id} = 0] - \mathbb{E}[Y_{id}(0)|\Delta_{id} = 0] \\
&= \tau_{RD},
\end{aligned}
$$

where the first line follows by definition, the second line follows from Assumption 2 and the sum law of limits, the third line follows from the consistency of potential outcomes, the fourth line follows from Assumption 1, and the fifth line uses linearity of expectations to rewrite the expression as the causal effect of T_{id}. By the consistency of $\widehat{\tau}_{RD}$ and the bounded variance ensured by Assumption 2, $\widehat{\tau}_{RD}$ is asymptotically unbiased. ∎

Proof of Proposition 2: The following derivation proves the result:

$$
\begin{aligned}
\widehat{\tau}_{PCRD} &= \widehat{\mu}_+(0|X_{id} = 1, X_{jd} = 0) - \widehat{\mu}_+(0|X_{id} = 0, X_{jd} = 1) \\
&\xrightarrow{p} \lim_{v \downarrow 0} \mathbb{E}[Y_{id}|\Delta_{id} = v, X_{id} = 1, X_{jd} = 0] - \lim_{v \downarrow 0} \mathbb{E}[Y_{id}|\Delta_{id} = v, X_{id} = 0, X_{jd} = 1] \\
&= \mathbb{E}[Y_{id}(1)|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] - \mathbb{E}[Y_{id}(1)|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1] \\
&= \mathbb{E}[\tau_d + g(\mathbf{Z}_{id}) + \upsilon_d|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] - \mathbb{E}[g(\mathbf{Z}_{id}) + \upsilon_d|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1] \\
&= \mathbb{E}[\tau_d|\Delta_d = 0] + \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] - \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1] \\
&= \tau_{PCRD} + \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] - \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1],
\end{aligned}
$$

where the first line follows from equation (4), the second line follows from Assumption 2 and the sum law of limits, the third line follows from the consistency of potential outcomes and Assumption

1(a), the fourth line substitutes the structural definition of potential outcomes in Assumption 4, the fifth line uses the linearity of the expectation operator and independence of v_d from all other variables, and the sixth line follows from $\tau_{PCRD} = \mathbb{E}[Y_d(1) - Y_d(0)|\Delta_d = 0] = \mathbb{E}[\tau_d|\Delta_d = 0]$. ∎

Proof of Proposition 3: For the identification result, the estimand can be rewritten as follows:

$$
\begin{aligned}
\mathbb{E}[Y_d(1) - Y_d(0)|\Delta_d = 0] &= \mathbb{E}[\tau_d|\Delta_d = 0] \\
&= \mathbb{E}[\tau_d|\Delta_d = 0] + \mathbb{E}[g(\mathbf{Z}_{id}) + v_d|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] \\
&\quad - \mathbb{E}[g(\mathbf{Z}_{id}) + v_d|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] \\
&\quad + \mathbb{E}[g(\mathbf{Z}_{id}) + v_d|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1] - \mathbb{E}[g(\mathbf{Z}_{id}) + v_d|\Delta_{id} = 0, X_{id} = 0, X_{jd} \\
&= \mathbb{E}[\tau_d + g(\mathbf{Z}_{id}) + v_d|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] \\
&\quad - \mathbb{E}[g(\mathbf{Z}_{id}) + v_d|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1] \\
&\quad - \left(\mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] - \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1] \right) \\
&= \mathbb{E}[Y_{id}(1)|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] \\
&\quad - \mathbb{E}[Y_{id}(1)|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1] - b_{PCRD} \\
&= \lim_{v \downarrow 0} \mathbb{E}[Y_{id}|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] \\
&\quad - \lim_{v \downarrow 0} \mathbb{E}[Y_{id}|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1] - b_{PCRD},
\end{aligned}
$$

where the first line follows by definition, the second line adds and subtracts additional terms, the third line combines terms using the independence of v_d given in Assumption 3, the fourth line applies the definition from Assumption 3 and denotes $b_{PCRD} := \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] - \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1] = 0$, and the final line uses the consistency of potential outcomes and applies Assumption 1(a).

To demonstrate sufficiency for consistency and identification, it suffices to show that each condition in Proposition 2 implies that $b_{PCRD} = 0$, which is the bias term in the proposition.

For condition (i): the following derivation establishes the result:

$$
\begin{aligned}
b_{PCRD} &= \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id}(X_{id},X_{jd},\mathbf{Z}_{id},\mathbf{Z}_{jd},\varepsilon_{id},\varepsilon_{jd}) = 0, X_{id} = 1, X_{jd} = 0] \\
&\quad - \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id}(X_{id},X_{jd},\mathbf{Z}_{id},\mathbf{Z}_{jd},\varepsilon_{id},\varepsilon_{jd}) = 0, X_{id} = 0, X_{jd} = 1] \\
&= \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id}(\mathbf{Z}_{id},\mathbf{Z}_{jd},\varepsilon_{id},\varepsilon_{jd}) = 0, X_{id} = 1, X_{jd} = 0] \\
&\quad - \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id}(\mathbf{Z}_{id},\mathbf{Z}_{jd},\varepsilon_{id},\varepsilon_{jd}) = 0, X_{id} = 0, X_{jd} = 1] \\
&= \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id}(\mathbf{Z}_{id},\mathbf{Z}_{jd},\varepsilon_{id},\varepsilon_{jd}) = 0] - \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id}(\mathbf{Z}_{id},\mathbf{Z}_{jd},\varepsilon_{id},\varepsilon_{jd}) = 0] \\
&= 0
\end{aligned}
$$

where the first lines rewrites Δ_{id} in terms of its arguments, the second line applies the independence of V_{id} and X_{id} and X_{jd} given in condition (i), and the third line follows from Assumption 3.

For condition (ii): the condition immediately yields $b_{PCRD} = 0$.

For condition (iii): the condition implies that $\mathbf{Z}_{id} = \mathbf{Z}_{jd}$ whenever $\Delta_{id} = 0$. This implies that $\mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] = \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1]$ whenever $\Delta_{id} = 0$, and thus $b_{PCRD} = 0$. ∎

Proof of Proposition 4: The following derivation proves the result:

$$
\begin{aligned}
\widehat{\tau}_{PCRD} &= \widehat{\mu}_+(0|X_{id} = 1, X_{jd} = 0) - \widehat{\mu}_+(0|X_{id} = 0, X_{jd} = 1) \\
&\overset{p}{\to} \lim_{v \downarrow 0} \mathbb{E}[Y_{id}|\Delta_{id} = v, X_{id} = 1, X_{jd} = 0] - \lim_{v \downarrow 0} \mathbb{E}[Y_{id}|\Delta_{id} = v, X_{id} = 0, X_{jd} = 1] \\
&= \mathbb{E}[Y_{id}(1)|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] - \mathbb{E}[Y_{id}(1)|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1] \\
&= \mathbb{E}[y(1,\mathbf{Z}_{id}) + v_d|\Delta_{id} = 0, X_{id} = 1, X_{jd} = 0] - \mathbb{E}[y(0,\mathbf{Z}_{id}) + v_d|\Delta_{id} = 0, X_{id} = 0, X_{jd} = 1] \\
&= \int y(1,\mathbf{z}) f_{\mathbf{Z}_{id}|\Delta_{id}=0,X_{id}=1,X_{jd}=0}(\mathbf{z}) d\mathbf{z} - \int y(0,\mathbf{z}) f_{\mathbf{Z}_{id}|\Delta_{id}=0,X_{id}=0,X_{jd}=1}(\mathbf{z}) d\mathbf{z} \\
&= \int [y(1,\mathbf{z}) - y(0,\mathbf{z})] f_c(\mathbf{z}) d\mathbf{z} + \int y(1,\mathbf{z}) f_1(\mathbf{z}) d\mathbf{z} - \int y(0,\mathbf{z}) f_0(\mathbf{z}) d\mathbf{z},
\end{aligned}
$$

where the first line follows from equation (4), the second line follows from Assumption 2 and the sum law of limits, the third line follows from the consistency of potential outcomes and Assumption

A5

1(a), the fourth line substitutes the structural definition of potential outcomes, the fifth line applies the definition of conditional expectation and the independence of υ_d from all other variables, and the final line uses the definitions of $f_c(\mathbf{z})$, $f_1(\mathbf{z})$, and $f_0(\mathbf{z})$ given in the proposition. ■

A.3 Example with two compensating differentials

While the main paper considered the simplest example with a single compensating differential, similar insights obtain for multiple compensating differentials. Consider a case with three relevant candidate characteristics: the characteristic of interest, $X_{id} \in \{0,1\}$; and two compensating differentials, $\mathbf{Z}_{id} = (Z_{id1}, Z_{id2})$. Two compensating differentials facilitates study of the general case where compensating differentials can be correlated and act in opposing ways.

I assume that $Z_{id1} - Z_{jd2}$ and $Z_{id2} - Z_{jd2}$ are both normally distributed across candidates according to $N(0,1)$, with a correlation between these differences of $\rho \in (-1,1)$. By normalizing the standard deviations of the differences in compensating differentials to 1, the effects of each on vote share—β_k and δ_k—can be interpreted as standardized effects. Following Assumption 3, both $Z_{id1} - Z_{jd2}$ and $Z_{id2} - Z_{jd2}$ are independent of X_{id}. The difference between the idiosyncratic vote share shocks $\varepsilon_{id} - \varepsilon_{jd}$ is independently distributed across districts according to $N(0, \sigma_\varepsilon^2)$. I impose the following functional forms on the vote and outcome functions:

$$
\begin{aligned}
V_{id} &= \alpha \frac{X_{id} - X_{jd}}{2} + \sum_{k=1,2} \beta_k \frac{Z_{idk} - Z_{jdk}}{2} + \frac{\varepsilon_{id} - \varepsilon_{jd}}{2}, \\
Y_{id}(1) &= \tau X_{id} + \sum_{k=1,2} \gamma_k Z_{idk} + \upsilon_d,
\end{aligned}
$$

where $\alpha \geq 0$ and $\beta_1, \beta_2 > 0$, and the outcome equation satisfies Assumption 4. Again, I slightly abuse notation because V_{id} is bounded and the distributions are not. However, for sufficiently small β_k and σ_ε^2, the bounds are almost never violated.

The following result characterizes the asymptotic bias of the PCRD estimator when there are two correlated compensating differentials:

Proposition A1. *Assume Assumptions 1(a)-4 hold. Where $Z_{idk} - Z_{jdk} \sim N(0,1)$ for each $k = 1,2$,*

$Corr[Z_{id1} - Z_{jd1}, Z_{id2} - Z_{jd2}] = \rho$, *and* $\varepsilon_{id} - \varepsilon_{jd} \sim N(0, \sigma_\varepsilon^2)$, *the PCRD estimator satisfies:*

$$\widehat{\tau}_{PCRD} \xrightarrow{p} \tau_{PCRD} - \sum_{k=1,2} \gamma_k \left(\frac{\alpha(\beta_k + \rho\beta_{-k})}{\beta_1^2 + \beta_2^2 + 2\rho\beta_1\beta_2 + \sigma_\varepsilon^2} \right),$$

where $-k$ *denotes the compensating differential that is not* k. *If* ρ *is not too negative,* $\widehat{\tau}_{PCRD}$ *converges to an underestimate (overestimate) in magnitude of* X_{id}'s *effect when the sign of* τ *agrees (disagrees) with the signs of* γ_1 *and* γ_2. *If* ρ *is sufficiently negative,* $\widehat{\tau}_{PCRD}$ *converges to an overestimate in magnitude of* X_{id}'s *effect when* $|\gamma_1 - \gamma_2|$ *is sufficiently large.*

Proof of Proposition A1. Given that $Z_{id1} - Z_{jd1}$, $Z_{id2} - Z_{jd2}$, and $\varepsilon_{jd} - \varepsilon_{id}$ are normally distributed as stated in the proposition:

$$\mathbb{E}[Z_{idk} - Z_{jdk}|\Delta_d] = \mathbb{E}[Z_{idk} - Z_{jdk}] + \frac{Cov[Z_{idk} - Z_{jdk}, \Delta_d]}{\mathbb{V}[\Delta_d]}(\Delta_d - \mathbb{E}[\Delta_d]),$$

where $\mathbb{E}[Z_{idk} - Z_{jdk}] = 0$ (by its distribution) and $\mathbb{E}[\Delta_d] = \alpha$ (because $Z_{id1} - Z_{jd1}$, $Z_{id2} - Z_{jd2}$, and $\varepsilon_{jd} - \varepsilon_{id}$ are all centered on 0), and $\mathbb{V}[\Delta_d] = \beta_1^2 + \beta_2^2 + 2\rho\beta_1\beta_2 + \sigma_\varepsilon^2$ (due to the independence of $\varepsilon_{id} - \varepsilon_{jd}$ from Z_{idk} and the distributional assumptions on Z_{id}), recalling that $\Delta_d = \alpha + \sum_{k=1}^{2} \beta_k(Z_{1dk} - Z_{0dk}) + \varepsilon_{1d} - \varepsilon_{0d}$. We then obtain the following expressions by conditioning on $\Delta_d = 0$, $X_{idk} = 1$, and $X_{jdk} = 0$:

$$\mathbb{E}[Z_{1d1} - Z_{0d1}|\Delta_d = 0, X_{id} = 1, X_{jd} = 0] = -\frac{\alpha(\beta_1 + \beta_2\rho)}{\beta_1^2 + \beta_2^2 + 2\rho\beta_1\beta_2 + \sigma_\varepsilon^2},$$

$$\mathbb{E}[Z_{1d2} - Z_{0d2}|\Delta_d = 0, X_{id} = 1, X_{jd} = 0] = -\frac{\alpha(\beta_2 + \beta_1\rho)}{\beta_1^2 + \beta_2^2 + 2\rho\beta_1\beta_2 + \sigma_\varepsilon^2},$$

where the covariance terms reflect independence of $\varepsilon_{id} - \varepsilon_{jd}$ from $Z_{idk} - Z_{jdk}$ and the distributional assumptions on $Z_{idk} - Z_{jdk}$. Applying Proposition 2 then yields the result in Proposition A1.

It is obvious from inspection of the result in Proposition A1 that, when ρ is not too negative and $\tau > 0$, X_{id}'s effect is underestimated (overestimated) when $\gamma_1, \gamma_2 > (<)0$. The reverse holds when $\tau < 0$. When ρ is sufficiently negative and each γ_k shares the same sign ($\gamma_k \geq 0$, without loss of generality), it is possible that $\beta_1\gamma_1 + \beta_2\gamma_2 + \rho(\beta_1\gamma_2 + \beta_2\gamma_1) < 0$ (while the variance in the

denominator is always positive), and thus overestimation in magnitude occurs when τ shares the sign of γ_k. At the lower bound of $\rho = -1$, the condition holds when $(\beta_1 - \beta_2)(\gamma_1 - \gamma_2) < 0$. $\blacksquare$

The additional insight that emerges from this example with multiple compensating differentials is that $\widehat{\tau}_{PCRD}$ can also be consistent in the knife-edge case where conflicting effects of the compensating differentials on the outcome exactly cancel out at the point of discontinuity. Such a scenario would only hold if voters systematically failed to appreciate the implications of at least one compensating differential for district-level outcomes. For example, competent candidates might be less popular despite competence improving post-election outcomes in their district.

A.4 Do PCRD designs reduce bias relative to observational designs?

Since no bias mitigation strategy is perfect, it is natural to wonder whether the asymptotic biases introduced by PCRD designs exceed the biases of alternative research designs. To speak to this question, I compare the PCRD design to a difference in means design within the general framework of section 3.4. The difference in means estimator is given by $\widehat{\tau}_{PCDM} := \mathbb{E}[Y_{id}|\widehat{\Delta_{id} > 0}, X_{id} = 1] - \mathbb{E}[Y_{id}|\widehat{\Delta_{id} > 0}, X_{id} = 0]$, where this comparison of winners of different types need not restrict attention to races where the top two candidates differed in their type X_{id}. (More complex conditioning strategies, such as selection on observables or difference-in-differences, could be used instead.) This design, which includes observations from races that are not close, estimates effects of X_{id} among candidates i that won with any victory margin $\Delta_{id} > 0$.

The following proposition establishes when $\widehat{\tau}_{PCRD}$ is less asymptotically biased—relative to its target estimand—than $\widehat{\tau}_{PCDM}$ is relative to its distinct target estimand:

Proposition A2. *Under Assumptions 1(a), 2, and 4, $\widehat{\tau}_{PCRD}$ is less asymptotically biased in magnitude than $\widehat{\tau}_{PCDM}$ when:*

$$
\begin{aligned}
&\left| \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_d = 0, X_{id} = 1, X_{jd} = 0] - \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_d = 0, X_{id} = 0, X_{jd} = 1] \right| \\
&\leq \left| \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_d > 0] - \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_d \leq 0] \right|.
\end{aligned}
$$

Proof of Proposition A2: The magnitude of the bias on the left hand side follows directly from Proposition 2. The bias of $\widehat{\tau}_{PCDM}$ is derived as follows:

$$
\begin{aligned}
\widehat{\tau}_{PCDM} \xrightarrow{p}\ & \mathbb{E}[Y_d|\Delta_{id} > 0, X_{id} = 1] - \mathbb{E}[Y_d|\Delta_{id} > 0, X_{id} = 0] \\
=\ & \mathbb{E}[Y_{id}(1)|\Delta_{id} > 0, X_{id} = 1] - \mathbb{E}[Y_{id}(1)|\Delta_{id} > 0, X_{id} = 0] \\
=\ & \mathbb{E}[\tau_d|\Delta_{id} > 0] + \mathbb{E}[g(\mathbf{Z}_{id}) + \upsilon_d|\Delta_{id} > 0, X_{id} = 1] - \mathbb{E}[g(\mathbf{Z}_{id}) + \upsilon_d|\Delta_{id} > 0, X_{id} = 0] \\
=\ & \mathbb{E}[\tau_d|\Delta_{id} > 0] + \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} > 0, X_{id} = 1] - \mathbb{E}[g(\mathbf{Z}_{id})|\Delta_{id} > 0, X_{id} = 0],
\end{aligned}
$$

where the first line follows from consistent estimation of the conditional means under Assumption 2, the second follows from the consistency of individual-level potential outcomes, the third line follows from Assumption 4, and the fourth line follows from the independence of υ_d from all other variables. The bias of the difference in means estimator is $\mathbb{E}[\widehat{\tau}_{PCDM}] - \mathbb{E}[\tau_d|\Delta_{id} > 0]$. ∎

The inequality indicates that PCRD design are less biased when differences in outcomes due to differences in $\mathbf{Z}_{id}$ at the point of discontinuity are smaller than the corresponding differences between candidates of different types that won with any vote share. While it is plausible that candidates with the same vote share are more similar in terms of $\mathbf{Z}_{id}$, conditioning on close races accentuates the post-treatment bias because $\Delta_{id} = 0$ is a stricter condition than $\Delta_{id} > 0$. It is ultimately an application-specific empirical question which estimator is more biased relative to its estimand. However, it should be noted that $\widehat{\tau}_{PCDM}$ is likely to yield more precise estimates due to the larger sample size that such an analysis permits.

A.5 Differential empirical implications of theorized mechanisms

In addition to bounding exercises, researchers can attempt to separate compound treatments by testing implications that help to distinguish which candidate characteristics drive PCRD estimates. The goal of this approach is to substantiate a form of the claim that compensating differentials do not affect the outcome of interest, i.e. condition (ii) in Proposition 3. Observing meaningful discontinuities among other variables may again be useful in guiding the components of the com-

pound treatment that researchers should focus attention on. It should be emphasized that analyses of mechanisms rely on strong assumptions (Bullock, Green and Ha 2010), and are more plausible when there are few compensating differentials to distinguish the effects of characteristic X_{id} from. In practice, however, it is hard to be certain what compensating differentials—or how many—exist.

Researchers could provide evidence consistent with some components of the compound treatment driving an effect, and not others, in at least two ways. A first approach uses theory to identify post-treatment variables that are expected to change if X_{id} drives the effect at the discontinuity, but are not expected to change if relevant compensating differentials instead drive the effect. For example, Gagliarducci and Paserman (2012) argue that municipal governments led by women are more likely to terminate early due to resistance from lower-ranked men on the municipal council. If the PCRD estimate of the effect of gender on early termination were confounded by women who won close elections possessing different levels of competence than men who won close elections, then we should expect the treatment to influence government performance outcomes as well. The lack of such evidence in Gagliarducci and Paserman (2012) suggests that competence is not driving the PCRD estimates, either because it is not a compensating differential or because competence does not affect local government outcomes in the study's specific context.

A second approach—based on moderation—leverages subgroup variation where there are theoretical or contextual reasons to believe that some components of a compound treatment are more likely to be activated in certain subgroups than other components (e.g. Eggers et al. 2018). For example, Gagliarducci and Paserman (2012) argue that the effect of electing a woman on early government termination should be greater in parts of Italy where unfavorable attitudes toward women are more prevalent and there was limited prior history of women in office, whereas the downstream effect of electing a more competent candidate may not vary with such baseline conditions. It should be cautioned that causally attributing differences across subgroups to a particular covariate also requires exogenous variation in that covariate.