Can close election regression discontinuity designs identify effects of winning politician characteristics?

John Marshall
Columbia University

Regression discontinuity designs

Exploit a known threshold in a "forcing" variable to assign treatment

- E.g. vote margin determines election victory

Identification relies on a weak continuity assumption (see Cattaneo and Titiunik 2022)

Often regarded as the most credible observational identification strategy (Lee 2008), albeit with limited external validity

Now common in the social science articles and training (de la Cuesta and Imai 2016)

Two types of RD design using close elections

Standard RDs compare **narrow winners with narrow losers**

- Returns to office for candidates (e.g. Eggers and Hainmueller 2009)
- Party incumbency advantage, where parties always run (e.g. Lee et al. 2004)

Politician characteristic RD (PCRD) designs compare **narrow winners of different types** in races between different types

- Ascriptive characteristics, e.g. gender, race, clan, religion
- Prior actions of a politician, e.g. incumbency, corruption, education, criminality, vocation
- Labels politicians sort into, e.g. party, ideology
- Institutional status, e.g. partisan alignment across levels of government, term limited, seniority

PCRD designs

The appeal:

- Elected/selected politician characteristics may matter for downstream elections, government responsiveness, and citizen welfare and political participation
- Close elections between type $X = 1$ or $X = 0$ as-if randomly vary which type wins

Leading estimands of interest to analysts:

- Isolate the effect of X (possibly a bundle), holding other politician characteristics $\mathbf{Z}$ equal
- Effect of the bundle of *all* politician-level characteristics that come with $X = 1$ relative to $X = 0$ (in close elections)

...I largely focus on the former

126 published papers using PCRD designs; >11,000 cites

		...of which, employ/demonstrate awareness of...			
	Number of articles	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

Two PCRD-specific identification challenges

Correlated characteristics: close elections do not exogenously vary X (e.g. Sekhon and Titiunik 2012)

- E.g. women that narrowly win remain more likely to be Democrats
- Acknowledged by 33% of studies

Compensating differentials: even if X was independent of all $\mathbf{Z}$s, conditioning on close elections can induce post-treatment bias

- E.g. women in close races are more likely to be competent
- Only even loosely recognized by 6% of studies

This paper

Connects PCRD to standard RD designs to distinguish them

Identifies two general conditions under which post-treatment bias arises/additional assumptions required to isolate the LATE of X:

- Characteristic X affects vote share V
- Compensating differentials $\mathbf{Z}$ affect outcome Y

Characterizes the direction of the asymptotic bias

Implications for best practice:

- Can we credibly invoke the additional assumptions?
- What can be done without imposing additional assumptions?
- Should we redefine the estimand to capture a treatment that bundles all politician characteristics that come with X?

A more fundamental challenge to PCRD designs

Other ways compound treatments can emerge:

- Multiple treatments assigned together (Eggers et al. 2018)
- Treatment alters *subsequent* comparisons (Eggers 2017; Sekhon and Titiunik 2012)
- Dispute over whether close elections satisfy continuity (Caughey and Sekhon 2011 vs. Eggers et al. 2015)

Estimation issues:

- Misspecification bias and control function selection (Cattaneo and Titiunik 2022; Gelman and Imbens 2019)
- Limited statistical power (Stommes et al. 2023)
- Bandwidth selection (Imbens and Kalyanaraman 2012)

...bias remains when continuity holds and absent estimation issues

- Importance of theory for empirical models (e.g. Ashworth et al. 2021; Bueno de Mesquita and Tyson 2020; Eggers 2017; Slough forthcoming)

A motivating example: politician gender

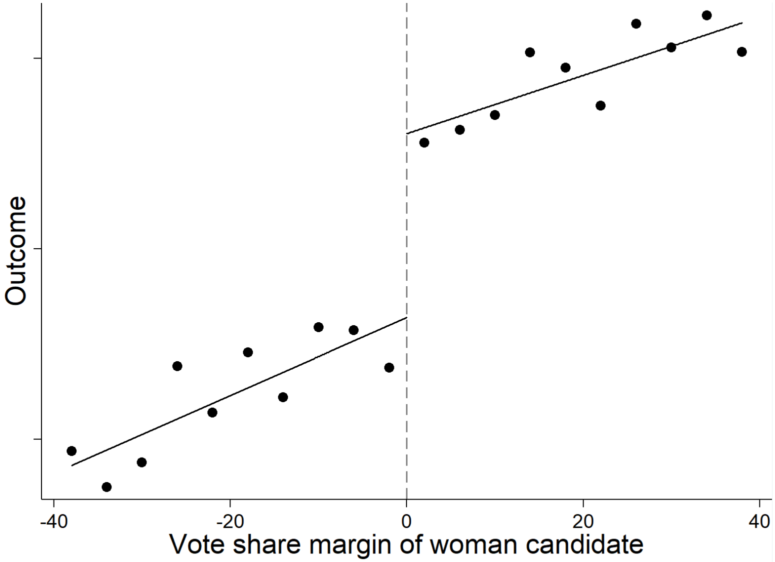
Estimand: effect of politician gender on district employment levels

Define binary treatment X as distinct from other characteristics $\mathbf{Z}$

- Assume competence is regarded as distinct from gender by the researcher
- But some candidate attributes are immutable or inherently bundled, other analysts may care about the overall bundle rather than specific characteristics

Sample: districts where women just beat men or vice versa

What we estimate



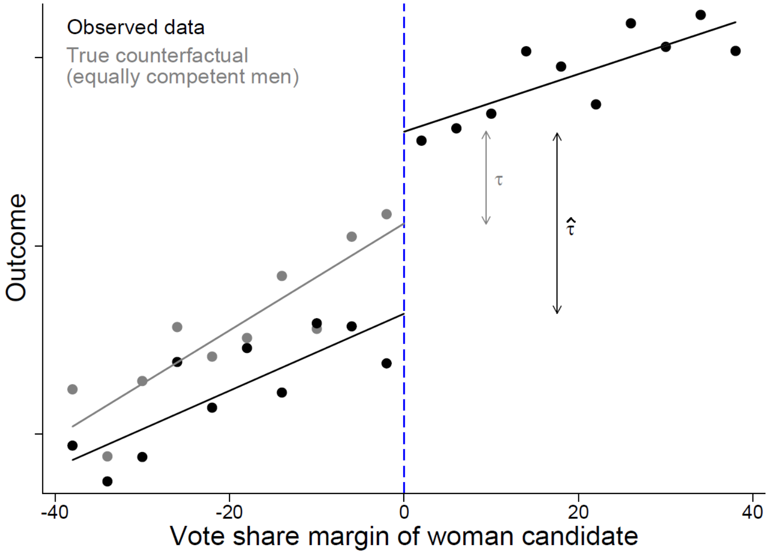
Compensating differentials

But are women in close races similar to men in close races?

- Anti-woman biases among voters (e.g. Lawless 2015)
- ...women in close races must be more competent on average

Compensating differentials: the vector of characteristics $\mathbf{Z}$ that ensures candidates with characteristic X are in close races with candidates without X, which may also affect outcome Y

Upward bias due to competence



Another example

Estimand: effect of politician education on district employment levels

Compensating differential: more effective policies

Effect of education is downwardly biased if:

- More educated candidates are electorally advantaged in general; and
- Greater education and better policies produce more employment

What's different about PCRD designs?

Focus on single-member district plurality election

- Similar logic applies to threshold representation, legislative majorities, last person elected to multi-member districts

Let's formally connect standard RD and PCRD designs...

Standard regression discontinuity designs

Candidate i in district d receives share $V_{id} \in [0,1]$ of top 2 votes

Forcing variable: $\Delta_{id} = V_{id} - V_{jd} \in [-1,1]$, where $j \neq i$ is the most popular candidate other than i

Election win treatment variable:

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

Assumption 1 (continuity): Potential outcomes $Y_{id}(T_{id}) \in \mathbb{R}$ satisfy:

(a) $\lim_{v \downarrow 0} \mathbb{E}[Y_{id}(1) | \Delta_{id} = v] = \mathbb{E}[Y_{id}(1) | \Delta_{id} = 0]$
(b) $\lim_{v \uparrow 0} \mathbb{E}[Y_{id}(0) | \Delta_{id} = v] = \mathbb{E}[Y_{id}(0) | \Delta_{id} = 0]$

Consistency: $Y_{id} = T_{id} Y_{id}(1) + (1 - T_{id}) Y_{id}(0)$

Identification of LATE

Target estimand: $\tau_{RD} = \mathbb{E}[Y_{id}(T_{id} = 1) - Y_{id}(T_{id} = 0) | \Delta_{id} = 0]$

Estimator in a sample of n elections (from a superpopulation):

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

where $\widehat{\mu}_{+}(s)$ estimates $\lim_{v \downarrow s} \mathbb{E}[Y_{id} | \Delta_{id} = v]$ and $\widehat{\mu}_{-}(s)$ estimates $\lim_{v \uparrow s} \mathbb{E}[Y_{id} | \Delta_{id} = v]$

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

Proposition 1 (Hahn et al. 2001): Under Assumptions 1 and 2, $\widehat{\tau}_{RD}$ is a consistent and asymptotically unbiased estimator of τ_{RD}

How PCRD designs differ

Unit of analysis is now the district, not the candidate:

- Forcing variable: $\Delta_d = V_{1d} - V_{0d}$, where 1 and 0 denote the most popular politicians of type $X_{id} = 1$ and $X_{id} = 0$ in d
- Treatment: $X_d = \begin{cases} 1 & \text{if } \Delta_d > 0 \\ 0 & \text{if } \Delta_d \leq 0 \end{cases}$
- Potential outcomes: $Y_d(X_d) = X_d\, Y_{1d}(1) + (1 - X_d)\, Y_{0d}(1)$

Target estimand that isolates the effect of X_d from $\mathbf{Z}_d$:

$$\tau_{PCRD} = \mathbb{E}[Y_d(X_d = 1) - Y_d(X_d = 0) | \Delta_d = 0]$$

Estimator:

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

PCRD designs differ by *focusing only on winners and conditioning on a predetermined difference in X_{id} that could affect Δ_d*

Defining what is and is not part of treatment

Often conceptually challenging to define X_{id}

- Are competence differences compensating differentials or part of the gender effect?
- Are differences in policies compensating differentials or part of the education effect?

Start by defining an X_{id} that is predetermined wrt election outcomes and is conceptually distinct from at least some $\mathbf{Z}_{id}$

Assumption 3 (independence): X_{id} is independent of $\mathbf{Z}_{id}$ and $\mathbf{Z}_{jd}$ (at least among politicians that could enter close elections)

...assumes away correlated characteristics problem, for a "best case scenario" for PCRD designs seeking to isolate the effect of X

Example with a single compensating differential

Single compensating differential $Z_{id} - Z_{jd} \sim N(0, \sigma_Z^2)$, where Z_{id} and Z_{jd} are independent of X_{id}

Candidate vote share:

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

where $\alpha \geq 0$ and $\beta > 0$, and $\varepsilon_{id} - \varepsilon_{jd} \sim N(0, \sigma_\varepsilon^2)$ is noise

Potential outcomes with constant effects:

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

where υ_d is noise that is independent of all variables

The close election condition

In the limiting case of close elections:

$$V_{1d} = V_{0d} \iff \alpha + \beta(Z_{1d} - Z_{0d}) + \varepsilon_{1d} - \varepsilon_{0d} = 0$$

If $\alpha > 0$, then an election is close because:

- Compensating differentials exist: $Z_{1d} < Z_{0d}$
- Noise favored the $X_{id} = 0$ candidate: $\varepsilon_{1d} < \varepsilon_{0d}$

Asymptotic bias of the PCRD estimator

Under Assumptions 1(a) and 2 and the distributional assumptions:

$$\widehat{\tau}_{PCRD} \xrightarrow{p} \tau - \gamma \frac{\alpha\beta \frac{\sigma_Z^2}{\sigma_\varepsilon^2}}{1 + \beta^2 \frac{\sigma_Z^2}{\sigma_\varepsilon^2}}$$

Special cases of zero bias:

- No compensating differential is required ($\alpha = 0$)
- Compensating differential doesn't affect the outcome ($\gamma = 0$)
- Close elections only arise due to noise ($\alpha + \varepsilon_{1d} - \varepsilon_{0d} = 0, \forall d$)

Downward (upward) bias when sign of τ agrees (disagrees) with γ

- Examples: upward bias for gender, downward for education

Magnitude of bias increases with α, γ, and $\sigma_Z^2/\sigma_\varepsilon^2$

General result

Now consider an unrestricted vote share function $v(X_{id}, X_{jd}, \mathbf{Z}_{id}, \mathbf{Z}_{jd}, \varepsilon_{id}, \varepsilon_{jd})$

Assumption 4 (additive separability): Candidate i's potential outcome if elected is given by $Y_{id}(1) = \tau_d X_{id} + g(\mathbf{Z}_{id}) + \upsilon_d$

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

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

Bias reflects correlated characteristics or compensating differentials

General conditions for avoiding bias

Imposing Assumption 3 ensures bias only reflects post-treatment conditioning

Proposition 3: Under Assumptions 1-4, $\widehat{\tau}_{PCRD} \xrightarrow{p} \tau_{PCRD}$ 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 $v(1, 0, \mathbf{z}, \mathbf{z}', \varepsilon_{id}, \varepsilon_{jd}) = v(0, 1, \mathbf{z}', \mathbf{z}, \varepsilon_{jd}, \varepsilon_{id})$, $\mathbf{z} = \mathbf{z}'$

Strategies for isolating the effect of X

1. Invoking an additional assumption
2. Mitigating biases

Invoking an additional assumption

Impose condition (i) or (ii), given the implausibility of (iii)

Empirical limitations:

- For condition (ii): hard to measure all $\mathbf{Z}$ *and* show that none affects Y in close races
- For condition (i): more feasible to show that X doesn't affect V in races that end up being close

Theoretical limitation:

- If $\widehat{\tau}_{PCRD} \neq 0$, voters must be oblivious to X's impact on outcomes, not care about such outcomes, or care about other outcomes that net out
- This is challenging, since even irrational voters have beliefs about the desirability of X

Mitigating biases: continuity/balance tests

Conditions (i) and (iii) $\Rightarrow$ continuity tests operate as normal

Otherwise, PCRD designs *require* imbalances for some Z_{idk} (to satisfy the close election condition):

- Applies to individual, *not district*, level covariates
- Detecting a theoretically-plausible imbalance is now a "manipulation check" for PCRD designs
- Failing to detect imbalances is consistent with condition (i) or condition (iii) holding *and* failing to measure compensating differentials, a lack of statistical power, a false positive, etc. $\Rightarrow$ continuity tests are less informative

Mitigating biases: covariate adjustment

Adjust for observable compensating differentials $\mathbf{Z}_{id}^{cond} \subset \mathbf{Z}_{id}$

- A widespread strategy in practice

Limitations:

- Adjustment cannot solve post-treatment bias by making condition (i) more plausible; e.g. conditional on competence, women in these subsets of close races must still differ in another way
- But adjustment could increase the plausibility of conditions (ii) and (iii) by reducing $\frac{\sigma^2_{Z|\mathbf{Z}^{cond}}}{\sigma^2_{\varepsilon}}$

Mitigating biases: bounding

Information about effects of $\mathbf{Z}_{id}$ aids inference:

- Abandoning point estimates:
  - If net effect of $\mathbf{Z}_{id}$ agrees with X_{id}, $\widehat{\tau}_{PCRD}$ is an underestimate $\Rightarrow$ rejections of the null hypothesis are valid, but increases false negatives
  - If net effect of $\mathbf{Z}_{id}$ disagrees with X_{id}, $\widehat{\tau}_{PCRD}$ is an overestimate $\Rightarrow$ increases false positives
- Possible to combine (dis)continuity tests with sensitivity analyses, where $g(\mathbf{Z}_{id})$ is linear in each Z_{idk}:
  - Use (dis)continuity tests to estimate difference in each Z_{idk} at the discontinuity.
  - Impute possible effects of each Z_{idk} on Y_{id}

Limitations:

- Permits more limited claims
- Unobserved compensating differentials remain unaccounted for

Better PCRD than nothing?

Bias applies to any design that conditions the sample on winners

- Selection on observables, diff-in-diff

The PCRD trade-off:

- Close races may reduce correlations between X_{id} and $\mathbf{Z}_{id}$ by conditioning on V_{id} (essentially matching on a post-treatment covariate)
- Conditioning on $V_{id} = V_{ij}$ could increase post-treatment bias (and reduce power)

Expanding the conception of treatment

Treatment now constitutes the characteristic of interest *and all other correlated characteristics* in close elections

- E.g. focus on ideological extremists rather than ideological extremism (Hall 2015)
- (Dis)continuity tests help to *interpret* treatment

Express potential outcomes as $Y_d(X_d, \mathbf{Z}_d)$ to highlight how this alternative estimand differs:

- Isolating the effect of X: $\mathbb{E}[Y_d(1, \mathbf{z}) - Y_d(0, \mathbf{z}) | \Delta_d = 0]$ (usually further integrating over $\mathbf{z}$)
- Bundled effect of characteristics that come with type X: $\mathbb{E}[Y_d(1, \mathbf{Z}_d(1)) - Y_d(0, \mathbf{Z}_d(0)) | \Delta_d = 0]$, where $\mathbf{Z}_d$ is random (so candidates in close races with $X = 1$ can have different permutations of $\mathbf{Z}(1)$)

What the bundled treatment captures

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

$$\widehat{\tau}_{PCRD} \overset{p}{\rightarrow} \int \Big[y(1,\mathbf{z}) - y(0,\mathbf{z})\Big] f_c(\mathbf{z}) d\mathbf{z} + \int y(1,\mathbf{z}) f_1(\mathbf{z}) d\mathbf{z} - \int y(0,\mathbf{z}) f_0(\mathbf{z}) d\mathbf{z},$$

where $f_{\cdot}(\mathbf{z})$ is the common or excess conditional density of $\mathbf{z}$ at $\Delta_d = 0$ in races between politicians of type $X_{id} = 1$ and $X_{jd} = 0$

Advantages:

- Identification doesn't on require strong additional assumptions
- The bundle may be a policy-relevant estimand

Limitations:

- Hard to characterize what is and is not part of treatment
- Lack of causal attribution limits testing of theories of X
- Heterogeneous bundles are "atypical," even of close races

Conclusions

Standard RD and PCRD designs differ in how they exploit the randomness of close elections

PCRD designs entail an identification-estimand tradeoff:

- Isolating effects of X requires assumptions far stronger than standard RD designs to avoid confounding by naturally correlated characteristics and design-induced compensating differentials
- Broadening treatment to include other politician characteristics $\mathbf{Z}$ captures a different estimand, which alters the theoretical or policy inferences that can be drawn

Regardless of the estimand selected, researchers should explicitly state their target estimand, conceptualization of treatment (especially what is *not* included), and assumptions required for identification