Political Analysis (2025), **0**, 1–8
doi:10.1017/pan.2024.28

PA

LETTER

When Can We Trust Regression Discontinuity Design Estimates from Close Elections? Evidence from Experimental Benchmarks

Leandro De Magalhães[1], Dominik Hangartner[2], Salomo Hirvonen[3], Jaakko Meriläinen[4], Nelson A. Ruiz[5] and Janne Tukiainen[3]

[1]Department of Economics, University of Bristol, Bristol, UK; [2]Center for Comparative and International Studies, ETH Zurich, Zurich, Switzerland and Department of Government, London School of Economics and Political Science, London, UK; [3]Department of Economics, University of Turku, Turku, Finland; [4]Department of Economics, Stockholm School of Economics, Stockholm, Sweden; [5]Department of Government, University of Essex, Colchester, UK

Corresponding author: Janne Tukiainen; Email: janne.tukiainen@utu.fi

(Received 6 March 2024; revised 1 July 2024; accepted 19 August 2024)

Abstract

Regression discontinuity designs (RDD) are widely used in the social sciences to estimate causal effects from observational data. Following recent methodological advances, scholars can choose from various RDD estimators for point estimation and inference. This decision is mainly guided by theoretical results on optimality and Monte Carlo simulations because of a paucity of research on the performance of the different estimators in recovering real-world experimental benchmarks. Leveraging exact ties in personal votes in local elections in Colombia and Finland, which are resolved by a random lottery, we assess the performance of various estimators featuring different polynomial degrees, bias-correction methods, optimal bandwidths, and approaches to statistical inference. Using re-running and re-election as outcomes, we document only minor differences in the performance of the various implementation approaches when the conditional expectation function (CEF) of the outcomes in the vicinity of the discontinuity is close to linear. When approximating the curvature of the CEF is more challenging, bias-corrected and robust inference with coverage-error-rate-optimal bandwidths comes closer to the experimental benchmark than more widely used alternative implementations.

Keywords: close elections; personal incumbency advantage; regression discontinuity design

Edited by: Jeff Gill

1. Introduction

Regression discontinuity designs (RDDs) are widely used in political science and neighboring disciplines. The popularity of RDD is not surprising given that it is often heralded as one of the few observational study designs that is successful in approximating experimental benchmarks (Green *et al.* 2009). The sharp RDD features a continuous forcing variable and a treatment assigned to units whose value of the forcing variable exceeds a known cutoff. Under the assumption that the conditional expectations of the potential outcomes are continuous in the forcing variable at the cutoff, RDD will, if correctly implemented, identify the average treatment effect at the cutoff. The success of RDD in recovering causal effects hinges on the precise approximation of the regression function above and below the cutoff, as the statistical properties of estimation and inference are closely tied to the accuracy of this approximation (Cattaneo, Idrobo, and Titiunik 2020). Point estimation of the treatment effect focuses on minimizing the mean squared error (MSE), a standard metric that evaluates the estimator's accuracy. In contrast, statistical inference aims to ensure that the empirical coverage of confidence intervals matches their nominal levels, which can be assessed using the coverage error rate (CER).

Researchers are required to make several critical decisions regarding the implementation of the RDD estimator. These decisions include selecting a bandwidth, determining a weighting scheme for observations near and far from the threshold, choosing the polynomial order for the locally weighted least squares regression, and deciding on a method for statistical inference (see, e.g., Cattaneo *et al.* 2020; Lee and Lemieux 2010). Approximating the conditional expectation function (CEF) of the outcomes in the vicinity of the discontinuity becomes more challenging with stronger curvature. Since different implementation methods vary in their capacity to capture this curvature with local polynomial functions, the choice of the implementation method becomes more relevant if the CEF is non-linear. The dominant approach has been to select the bandwidth by minimizing the MSE, use a rectangular or a triangular kernel to weight the sample, fit a linear polynomial on both sides of the threshold, and conduct inference using OLS approximations (Cattaneo *et al.* 2020; Lee and Lemieux 2010). We refer to this implementation as the "conventional" local linear estimation. While this point estimator is consistent and MSE-optimal, this approach leads to biased confidence intervals due to the approximation (smoothing) error of the local polynomial estimator (Calonico, Cattaneo, and Titiunik 2014; Cattaneo *et al.* 2020). Such inference is neither valid nor optimal in minimizing the CER. Frequently, researchers seek to correct this bias by using smaller than optimal bandwidths, combined with OLS approximation for inference (Lee and Lemieux 2010). However, such ad hoc undersmoothing leads to a loss of statistical power (Calonico *et al.* 2014; Calonico, Cattaneo, and Farrell 2020). The method proposed by Calonico *et al.* (2014) offers an alternative approach to inference by introducing bias-corrected and robust confidence intervals, circumventing the need for undersmoothing. This estimator first estimates the degree of bias using higher-order polynomials and then subtracts the estimated bias from the conventional point estimate. Robust inference is achieved by incorporating the contribution of the bias-correction step to the variability of the bias-corrected point estimator, thereby accounting for the estimation uncertainty of both the main RD estimate and the bias estimate. We refer to this implementation as the "robust" approach. Calonico *et al.* (2014) and Calonico *et al.* (2020) provide theoretical results and Monte Carlo simulations that suggest that the robust approach has lower coverage error rate than the conventional approach and ad hoc undersmoothing. Furthermore, Calonico, Cattaneo, and Farrell (2018) and Calonico *et al.* (2020) propose CER-optimal bandwidths that further improve the performance of the robust bias-corrected confidence intervals.

We reviewed 68 papers employing RDD that were published in the *American Political Science Review*, the *American Journal of Political Science*, and the *Journal of Politics* in 2016–2022 (see the Supplementary Material for an overview of the articles and implementations used). We found a rich tapestry of approaches to estimation and inference. The authors report local linear estimates in about nine out of ten articles, but MSE-optimal bandwidths are used in only about two-thirds of the published work. Conventional inference is most prevalent (in three out of four articles), and only 31% of the publications report the robust inference of Calonico *et al.* (2014). Maybe surprisingly, the use of robust inference has not increased over time (see Supplementary Figure OA1). The methodological variety of the surveyed literature further underscores the need for understanding the reliability and (relative) performance of different RDD approaches. While several studies examine the properties of different estimators and RDD implementations theoretically and through Monte Carlo simulations (Calonico *et al.* 2014, 2018, 2020), there has been little effort in evaluating their performance against experimental benchmarks.[1]

[1]Previous comparisons of RDD and experiments do not focus on implementation (Chaplin *et al.* 2018).

Our paper takes steps towards filling this evidence gap by documenting which RDD implementation can best replicate the gold standard of randomized experiments.[2]

Our validation analysis focuses on the electoral context, where numerous RDD applications have used close elections to estimate the effects of holding office on various outcomes, including incumbency advantage (for a review, see De la Cuesta and Imai 2016). Building on Hyytinen *et al.* (2018), we leverage electoral ties that are resolved by a lottery to estimate the effect of being the incumbent (versus being the runner-up) on re-running and getting elected in the next election in Colombia and Finland. Because candidates in tied elections have precisely the same number of votes, the average treatment effect estimated from the lottery sample is a local estimate at the cutoff that determines whether or not a candidate gets elected. This implies that in addition to focusing on the same institutional context and population, the lottery and RDD also target the same estimand. This makes lotteries an ideal benchmark to evaluate the performance of the RDD estimator. Both countries (mostly) use open-list proportional representation systems and provide us with a large number of observations to work with, even when we focus on close elections, making statistical power less of an issue (cf. Stommes *et al.* 2023). We extend on Hyytinen *et al.* (2018) by analyzing more countries and outcomes, which allows us to illuminate the role of the curvature of the outcome's CEF for the performance of various implementation approaches, and compare the results across different institutional contexts. Moreover, we include in our analysis the recent approach proposed by Calonico *et al.* (2020) who use for inference a bandwidth optimized for the CER (which is achieved by re-scaling the MSE-optimal bandwidth by a shrinkage factor proportional to the sample size)[3].

2. Data

Our main analysis examines local government elections in Colombia (2003—2015) and Finland (1996—2012).[4] Finland features a pure open-list electoral system where each voter gives exactly one vote to one candidate. Parties are assigned seats based on the sum of its candidates' personal votes, and the seats within the party are assigned purely on the basis of personal votes. Moreover, candidates are almost always presented in alphabetical order in the ballot lists. Council size depends on the municipal population and varies between 13 and 85. Councils are the main political decision-maker and are responsible for key public services such as education and healthcare. In Colombia, parties can choose between open or closed lists. However, in the 2015 local elections, about 92% of parties opted for open lists (Hangartner, Ruiz, and Tukiainen 2019), which are the focus of our analysis. Voters can still decide to vote just for the party, but personal votes determine the within-party allocation of seats. Council size varies between 7 and 45 and is determined by the number of registered voters. The main role of the council is to approve the budget and projects proposed by the mayor. In both countries, a sizable number of parties compete in local elections.

For the RDD analysis, we leverage party lists that nominate at least two candidates and elect at least one and fewer than all listed candidates. The resulting data consist of 147,558 candidate-election year observations for Colombia and 154,543 for Finland. The data reveal a substantial number of tied elections: 463 and 1,351 for Colombia and Finland, respectively. These samples are sufficiently large to provide reliable experimental benchmarks for comparing the RDD estimates. In the Supplementary Material, we show that there is no evidence of manipulation of the lottery outcomes.

[2]An understanding of which RDD implementation works best will also help reduce researchers' discretion in choosing among different implementations (Stommes, Aronow, and Sävje 2023).

[3]For the estimation of CER-optimal bandwidth, we use the rule-of-thumb implementation from Calonico *et al.* (2020), which is proportional to the robust bias-corrected bandwidth, and available in the rdrobust software package.

[4]The replication data and code available at Political Analysis Harvard Dataverse https://doi.org/10.7910/DVN/XDVIBG (De Magalhes *et al.* 2024).

3. Comparing Lottery and RDD Estimates

We first focus on the sample of tied candidates and regress, using OLS, an indicator variable for running or getting elected in the next election ($t+1$)—the two outcomes—on a binary indicator for getting elected in the current election (t)—the treatment. We do not condition the analysis of getting elected in $t+1$ on running in $t+1$ because this decision might be endogenous to getting elected in t.[5] We cluster our inference (and later optimal bandwidth selection) at the local government level.

Panel A of Table 1 shows the experimental estimates for running in the next election. Column (1) reports the effect in Colombia, and column (6) reports the effect in Finland. We find that in Colombia, getting elected boosts the probability of re-running by about 14 percentage points ($p = 0.002$). In the Finnish case, the point estimate is close to zero, about 0.011, and not statistically significant ($p = 0.671$). Columns (1) and (6) in Panel B of Table 1 show the estimation results for getting elected in the next election. The estimates for both Colombia and Finland are close to zero in magnitude, −0.030 and 0.004, and not statistically significant ($p = 0.371$ and $p = 0.860$, respectively). Thus, there is little evidence that being the winner in election t (versus being the runner-up) increases the probability of getting elected in $t+1$. Moreover, with 95% confidence intervals of $[-0.097, 0.037]$ for Colombia and $[-0.044, 0.053]$ for Finland, we can rule out all but relatively small incumbency effects.

We next turn to the RDD analysis. We construct the running variable from the winning margin for candidates on the same party list. For elected candidates, this equals their within-party vote share minus the within-party vote share of the first non-elected candidate. For the non-elected, this equals their within-party vote share minus the within-party vote share of the last elected candidate. This allows a comparison of candidates who barely won a seat to those who ran on the same list but barely lost. Columns (2)–(5) in Table 1 report the RDD estimates for Colombia, and columns (7)–(10) show the corresponding estimates for Finland. We provide eight specifications: conventional and robust approaches to inference, alternating between local linear and local quadratic polynomials, and using either MSE- or CER-optimal bandwidths. We use the same main and bias bandwidth for the robust bias-corrected estimation (Calonico *et al.* 2014, 2020). This means we effectively fit a polynomial of order $p+1$ within the bandwidth optimized for polynomial order p. The first implication arising from Panel A of Table 1 is that the lottery estimates for re-running are broadly in line with the RDD estimates for both countries. Although the lottery estimate is slightly smaller than the RDD estimates in the case of Colombia, these differences are not statistically significant. In the Finnish data, all differences between lottery and RDD estimates for re-running are miniscule.

When focusing on the incumbency advantage displayed in Panel B, we find larger deviations between the lottery and conventional RDD estimates. In both countries, we would draw qualitatively different conclusions regarding the incumbency effect estimated using the lotteries vis-à-vis RDD. While the lottery estimates provide little support for an incumbency advantage, the RDD estimates would imply a small positive and significant effect of getting elected in t on getting elected in $t+1$. However, when employing the robust RDD approach of Calonico *et al.* (2014), these discrepancies become more muted, especially when considering the uncertainty of the lottery and RDD estimates. Moreover, the CER optimal bandwidths yield confidence intervals that are closer to the experimental benchmark compared to the MSE-optimal bandwidths, even with the robust bias-corrected estimator. Lastly, we see that the CER optimal bandwidth used in combination with the conventional estimator does not solve the coverage issue as the bandwidth choice is optimized for the bias-corrected estimator.

The left graphs in Figure 1 visualize the RDDs. The plots show binned averages and local linear and quadratic fits within different (optimal) bandwidths. The right graphs plot the corresponding lottery estimates and RDD estimates using a range of bandwidths and local linear and local quadratic polynomials. Panel A shows the RDD for running in $t+1$ and Panel B for getting elected in $t+1$.

Panel A shows a positive RDD estimate of getting elected in t on the likelihood of running in $t+1$ in Colombia, with the size of the jump similar to the lottery estimate. For Finland, the RDD estimate is close

[5]However, the Supplementary Material provides additional results using election and vote share at $t+1$ conditional on re-running.

Table 1. Effect of incumbency on running in and winning the next election.

	Colombia					Finland				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Approach	Lotteries	RDD	RDD	RDD	RDD	Lotteries	RDD	RDD	RDD	RDD
Bandwidth selector		MSE	MSE	CER	CER		MSE	MSE	CER	CER
Polynomial		Linear	Quadratic	Linear	Quadratic		Linear	Quadratic	Linear	Quadratic
Panel A: Running t+1										
Lottery	0.141					0.011				
	[0.055,0.228]					[−0.040,0.062]				
Conventional		0.199	0.199	0.193	0.188		0.027	0.024	0.008	−0.000
		[0.178,0.220]	[0.178,0.221]	[0.169,0.217]	[0.163,0.212]		[0.004,0.049]	[0.002,0.046]	[−0.019,0.035]	[−0.027,0.026]
Robust		0.189	0.192	0.189	0.190		−0.004	−0.020	−0.024	−0.017
		[0.160,0.219]	[0.165,0.219]	[0.154,0.225]	[0.157,0.223]		[−0.040,0.032]	[−0.049,0.010]	[−0.063,0.015]	[−0.056,0.023]
N	463	45812	88944	31715	63086	1351	51357	100561	34401	79847
Bandwidth		4.38	9.75	2.98	6.28		1.48	3.50	1.05	2.37
Panel B: Elected t+1										
Lottery	−0.030					0.004				
	[−0.097,0.037]					[−0.044,0.053]				
Conventional		0.052	0.050	0.046	0.039		0.070	0.111	0.056	0.067
		[0.034,0.069]	[0.032,0.067]	[0.025,0.067]	[0.017,0.060]		[0.048,0.092]	[0.090,0.132]	[0.028,0.084]	[0.044,0.089]
Robust		0.040	0.034	0.037	0.042		0.022	0.041	−0.010	0.027
		[0.013,0.067]	[0.010,0.058]	[0.004,0.070]	[0.013,0.072]		[−0.016,0.060]	[0.016,0.066]	[−0.059,0.039]	[−0.003,0.058]
N	463	34526	72163	23618	49302	1351	30664	99582	20350	78807
Bandwidth		3.25	7.38	2.21	4.75		0.96	3.43	0.68	2.33

The dependent variable equals one if a candidate re-runs or gets elected in the next election and zero otherwise, in Panels A and B, respectively. Estimates in columns (1) and (6) are based on the election lottery samples. Columns (2)–(5) and (7)–(10) present results from different RDD specifications. “Conventional” refers to local linear estimation and OLS for inference. “Robust” refers to robust and biased-corrected inference and uses the main bandwidth for the bias-correction. All RDD estimations use a rectangular kernel. The 95% confidence intervals are based on standard errors clustered by municipality and reported in brackets. We also account for clustering when computing the optimal bandwidths. The number of observations refers to the effective sample size used for the estimation. The total number of observations is $147,558$ for Colombia and $154,543$ for Finland.

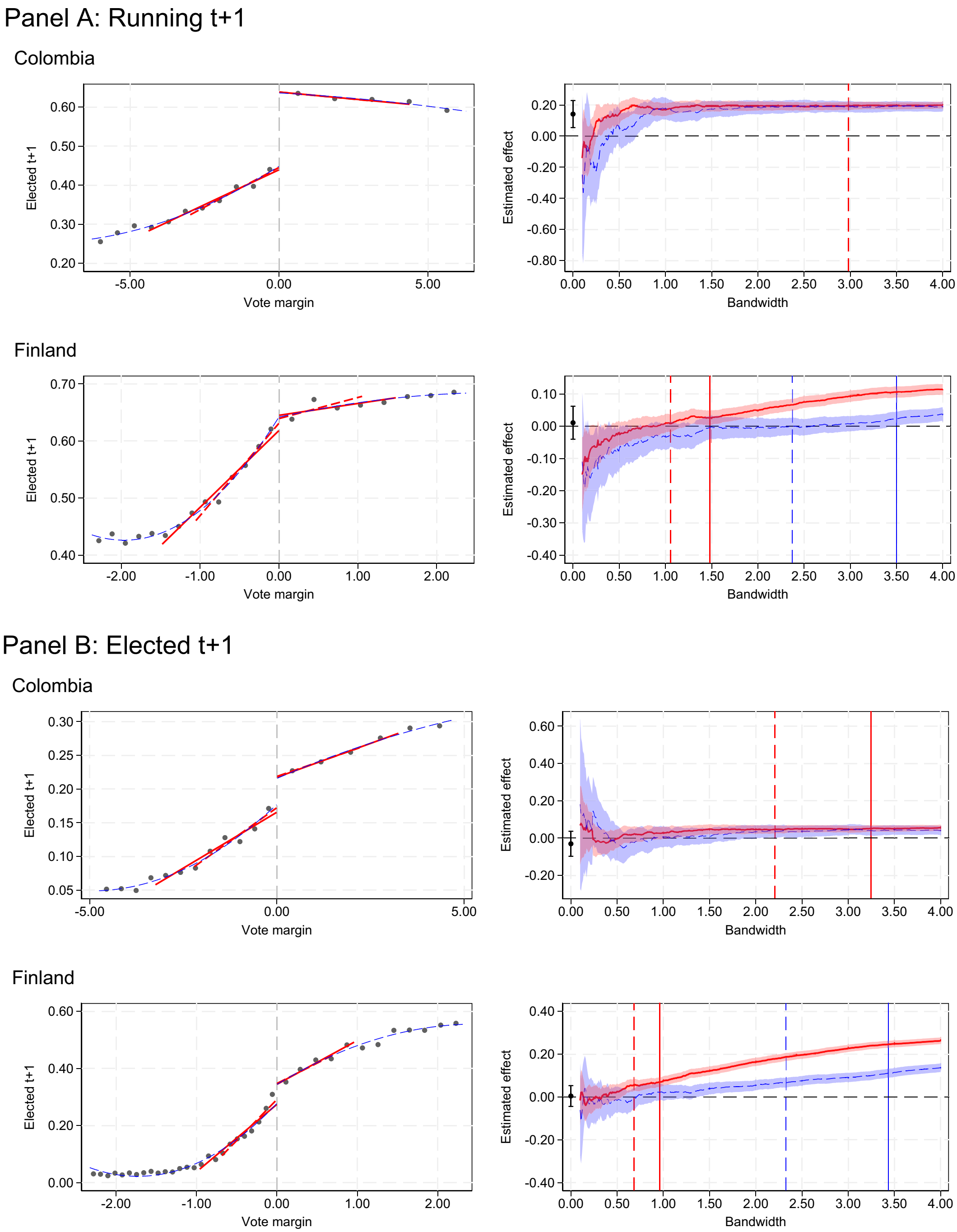

Figure 1. The figure shows RDD plots with binned averages (left) and RDD estimates across a range of bandwidths (right). The RDD plots on the left show local linear (red line) and local quadratic (blue line) fits within CER-optimal (dashed lines) and MSE-optimal (solid lines) bandwidths. The dependent variable is running in $t+1$ in Panel A and getting elected in $t+1$ in Panel B. The right plots show point estimates for the lottery sample (black) and the local linear (solid line) and local quadratic (dashed line) RDD specifications, obtained using a rectangular kernel. 95% confidence intervals are based on standard errors clustered at the municipality level. The dashed red and blue vertical lines indicate the CER-optimal bandwidths for the local linear and local quadratic estimation, respectively, and the solid vertical lines indicate the MSE-optimal bandwidths. For optimal bandwidths and corresponding point estimates and confidence intervals, see Table 1.

to zero, aligning with the experimental benchmark. In Panel B, the RDD estimates for the propensity to get elected in $t+1$ are positive for both Colombia and Finland, contrasting with the null findings from the lotteries.[6] However, in Finland, the graph suggests that the fitted polynomial models may inadequately capture the curvature of the CEF near the cutoff.[7] The right panel of Figure 1 demonstrates that discrepancies between the lottery estimates and the RDD graph can be mitigated by adjusting the approach to statistical inference; lower-order polynomials perform better in capturing curvature within narrower bandwidths, as seen in Panel B. Together, these results suggest an important finding: if the CEF is (approximately) linear close to the cut-off as in Panel A, both "conventional" and "robust" approaches can recover the experimental benchmark. If, however, the CEF is non-linear close to the cut-off as in Panel B, then the "robust" approach with CER-optimal bandwidths outperforms other implementations. In the next section, we discuss how this pattern extends to other data.

4. Discussion

Despite the popularity of RDD for drawing causal inferences from observational data, there is a paucity of research that evaluates if and when different RDD estimators are able to recover experimental estimates. Leveraging tied elections resolved by a lottery in Colombia and Finland as experimental benchmarks, we find that the type of RDD implementation makes little difference when the CEF around the cutoff is approximately linear. However, with curvature, the robust approach to inference proposed by Calonico *et al.* (2014) performs better than conventional local linear estimation. The linearity or non-linearity of the CEF close to the cutoff inform the choice of the bandwidth and in our applications, the CER methods suggests smaller bandwidths. This feature may explain why CER methods outperform MSE-optimal methods in this setting.

To understand whether the upward bias in the incumbency advantage documented for the conventional RDD estimate could be a symptom of a more widespread pattern, we extend our analysis to two neighboring countries with similar open-list PR systems: Brazil and Denmark. The Supplementary Material discusses the data and results in detail. In Brazil, we find relatively small differences between the different implementations when looking at running at $t+1$ as the dependent variable. In contrast, we again find larger estimates for the "conventional" compared to the "robust" approach in Denmark. Furthermore, the personal incumbency advantage estimates are smaller for both countries when we use the robust approach rather than conventional local linear estimation. A graphical analysis suggests that these differences are—again—partly due to the presence of curvature near the cutoff.

Our findings have both substantive and methodological implications. Substantively, our results suggest that the personal incumbency effect varies considerably across countries with similar electoral systems. Alternative explanations such as the weakness of the party system (Klašnja and Titiunik 2017) or the level of development and corruption (Klašnja 2015) are insufficient to explain the differences in incumbency advantage we observe across the four studied countries. Future research should explore other factors, such as differences in career objectives among politicians (De Magalhães and Hirvonen 2023), that may help explain the variation in incumbency advantage across these countries.

Methodologically, our study highlights the sensitivity of RDD estimates to specific implementation choices. The robust bias-corrected approach of Calonico *et al.* (2014) coupled with CER-optimal bandwidths proposed in Calonico *et al.* (2020) appears to (weakly) dominate other approaches, which is something we recommend practitioners keep in mind when using RDD. These implementation choices

[6]The Supplementary Material examines the robustness of the RDD estimates to alternative modeling choices. Specifically, it discusses how using separate optimal bandwidths for main estimation and bias correction can be beneficial when curvature near the cutoff is limited, but problematic when curvature is more pronounced. While kernel choice has minimal impact, the polynomial order significantly affects results. The Supplementary Material also tests robustness by controlling for incumbency, the most important predictor, and includes standard RDD validity checks.

[7]For non-linear CEFs, our analysis speaks in favor of CER-optimal bandwidths. Given that these bandwiths tend to be smaller than the conventionally used MSE-optimal ones, resulting in fewer effective observations, this might imply that even more RDD studies are underpowered than suggested by Stommes *et al.* (2023).

become more important when the curvature of the outcome's CEF close to the cutoff is not linear. We conclude by encouraging further research to better understand how the curvature of the outcome's CEF near the cutoff interacts with various RDD implementation choices, as this interaction plays a crucial role in the reliability of the RDD estimates.

Acknowledgements. We thank the editor, five anonymous referees, and conference participants at EPSA 2018 and the Econometric Society World Congress 2020 for their useful comments. We also thank Valtteri Kimmo for competent research assistance. This research is funded by the European Union (ERC, INTRAPOL, 101045239).

Funding. Views and opinions expressed are only those of the authors, however, and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Data Availability Statement. Replication code and data for this article are available online at Political Analysis Harvard Dataverse in De Magalhes *et al.* (2024), https://doi.org/10.7910/DVN/XDVIBG.

Supplementary Material. For supplementary material accompanying this paper, please visit https://doi.org/10.1017/pan.2024.28.

References

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2018. "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference." *Journal of the American Statictical Association* 113 (522): 767–779.

Calonico, S., M. D. Cattaneo, and M. H. Farrell. 2020. "Optimal Bandwidth Choice for Robust Bias-Corrected Inference in Regression Discontinuity Designs." *Econometrics Journal* 23 (2): 192–210.

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

Cattaneo, M., N. Idrobo, and R. Titiunik. 2020. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge Elements: Quantitative / Computational Methods for Social Science. Cambridge, MA: Cambridge University Press.

Chaplin, D. D., et al. 2018. "The Internal and External Validity of the Regression Discontinuity Design: A Meta-Analysis of 15 Within-Study Comparisons." *Journal of Policy Analysis and Management* 37 (2): 403–429.

De la Cuesta, B., and K. Imai. 2016. "Misunderstandings about the RD Design in the Study of Close Elections." *Annual Review of Political Science* 19: 375–396.

De Magalhães, L., and S. Hirvonen. 2023. "A Second Chance Elsewhere. Estimating the Effect of Winning (vs. Being the Runner-Up) on Future Electoral Prospects." *Electoral Studies* 83: 102612.

De Magalhes, L., D. Hangartner, S. Hirvonen, J. Meriläinen, N. A. Ruiz, and J. Tukiainen. 2024. "Replication Data for: When Can We Trust Regression Discontinuity Design Estimates from Close Elections? Evidence from Experimental Benchmarks." https://doi.org/10.7910/DVN/XDVIBG.

Green, D. P., T. Y. Leong, H. L. Kern, A. S. Gerber, and C. W. Larimer. 2009. "Testing the Accuracy of Regression Discontinuity Analysis Using Experimental Benchmarks." *Political Analysis* 17 (4): 400–417.

Hangartner, D., N. Ruiz, and J. Tukiainen. 2019. "Open or Closed? How List Type Affects Electoral Performance, Candidate Selection, and Campaign Effort." VATT Working Paper, 120.

Hyytinen, A., J. Meriläinen, T. Saarimaa, O. Toivanen, and J. Tukiainen. 2018. "When Does Regression Discontinuity Design Work? Evidence from Random Election Outcomes." *Quantitative Economics* 9 (2): 1019–1051.

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

Klašnja, M., and R. Titiunik. 2017. "The Incumbency Curse: Weak Parties, Term Limits, and Unfulfilled Accountability." *American Political Science Review* 111 (1): 129–148.

Lee, D. S., and T. Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355.

Stommes, D., P. M. Aronow, and F. Sävje. 2023." On the Reliability of Published Findings Using the Regression Discontinuity Design in Political Science." *Research & Politics* 10 (2): 20531680231166457.

Cite this article: De Magalhães, L., Hangartner, D., Hirvonen, S., Meriläinen, J., Ruiz, N. A. and Tukiainen, J., (2025). When Can We Trust Regression Discontinuity Design Estimates from Close Elections? Evidence from Experimental Benchmarks. *Political Analysis*, 1–8. https://doi.org/10.1017/pan.2024.28