

When does regression discontinuity design work? Evidence from random election outcomes

ARI HYYTINEN

School of Business and Economics, University of Jyväskylä

JAAKKO MERILÄINEN

Institute for International Economic Studies, Stockholm University

TUUKKA SAARIMAA

School of Business, Aalto University and VATT Institute for Economic Research

OTTO TOIVANEN

School of Business, Aalto University and KU Leuven

JANNE TUKIAINEN

Department of Government, London School of Economics and Political Science and
VATT Institute for Economic Research

We use elections data in which a large number of ties in vote counts between candidates are resolved via a lottery to study the personal incumbency advantage. We benchmark non-experimental regression discontinuity design (RDD) estimates against the estimate produced by this experiment that takes place exactly at the cutoff. The experimental estimate suggests that there is no personal incumbency advantage. In contrast, conventional local polynomial RDD estimates suggest a moderate and statistically significant effect. Bias-corrected RDD estimates that apply robust inference are, however, in line with the experimental estimate. Therefore, state-of-the-art implementation of RDD can meet the replication standard in the context of close elections.

KEYWORDS. Close elections, experiment, incumbency advantage, regression discontinuity design.

JEL CLASSIFICATION. C21, C52, D72.

Ari Hyttinen: ari.hyytinen@jyu.fi

Jaakko Meriläinen: jaakko.merilainen@iies.su.se

Tuukka Saarimaa: tuukka.saarimaa@vatt.fi

Otto Toivanen: otto.toivanen@aalto.fi

Janne Tukiainen: janne.tukiainen@vatt.fi

We thank anonymous referees, Manuel Bagues, Christine Benesch, Jon Fiva, Dominik Hangartner, Kaisa Kotakorpi, Benoit Le Maux, Yao Pan, Torsten Persson, Miikka Rokkanen, Riikka Savolainen, and conference and seminar participants at EPCS, EPSA, LSE, Rennes 1, Stockholm University, Warwick, and VATT for comments.

1. INTRODUCTION

A non-experimental empirical tool meets a very important quality standard if it can reproduce the results from a randomized experiment (LaLonde (1986), Fraker and Maynard (1987), Dehejia and Wahba (2002), and Smith and Todd (2005)). In a regression discontinuity design (RDD), individuals are assigned dichotomously to a treatment if they cross a given cutoff of an observable and continuous forcing variable, whereas those who fail to cross the cutoff form the control group (Thistlethwaite and Campbell (1960), Lee (2008), and Imbens and Lemieux (2008)). If the conditional expectation of the potential outcome is continuous in the forcing variable at the cutoff, correctly approximating the regression function above and below the cutoff, and comparing the values of the regression function for the treated and control groups at the cutoff gives the average treatment effect at the cutoff. We study whether RDD can, in practice, reproduce an experimental estimate that we obtain by utilizing data from electoral ties between two or more candidates in recent Finnish municipal elections.¹

The unique feature of our data is that ties were resolved via a lottery and that the random assignment occurs *right at the cutoff*. This feature means that if RDD works, it should produce an estimate that exactly matches our experimental estimate. Unlike in the prior work comparing RDD and an experiment, our experimental treatment effect is the same as the one that RDD targets. The setup of both the experiment and RDD refer to the same institutional context, to the same population of units, and basically to the same estimand.²

To explore whether RDD reproduces the experimental estimate, we utilize a data set that includes nearly 200,000 candidates who run for a seat in municipal councils in local Finnish elections every fourth year during 1996–2012. The elections were organized in a shared institutional environment and allow us to study whether there is a personal incumbency advantage, that is, extra electoral support that an incumbent politician of a given party enjoys when she runs for re-election, relative to her being a non-incumbent candidate from the same party and constituency (see, e.g., Erikson and Titiunik (2015)). Our experimental estimate of the personal incumbency advantage is estimated from data on 1351 candidates for whom the (previous) electoral outcome was determined via random seat assignment due to ties in vote counts.³ The experimental estimate provides

¹Investigating the performance of RDD in an electoral setting seems particularly important, as numerous applications of RDD have used close elections to estimate the effects of electoral results on a variety of economic and political outcomes (see, e.g., Lee, Moretti, and Butler (2004), Ferreira and Gyourko (2009), Gerber and Hopkins (2011), Folke and Snyder (2012), and De Magalhaes (2014)). de la Cuesta and Imai (2016) and Skovron and Titiunik (2015) are recent surveys of the close elections RDD analyses.

²Black, Galdo, and Smith (2007) come close to our analysis, because their experiment targets a population within a small bandwidth around the cutoff. However, as Black, Galdo, and Smith (2007, p. 107) point out, the experimental and non-experimental estimands are not quite the same in their setup: “Except in a common effect world, [...], the non-experimental estimators converge to a different treatment effect parameter than does the experimental estimator”.

³Use of lotteries to solve electoral ties is not unique to Finland. For example, some U.S. state elections and many U.S. local elections have used lottery-based rules to break ties in elections (see, e.g., UPI, July 14, 2014, *The Atlantic*, November 19, 2012, and Stone (2011)). Lotteries have been used to determine the winner in case of ties also in the Philippines (*Time*, May 17, 2013), in India (*The Telegraph India*, February

no evidence of a personal incumbency advantage; it is close to zero and quite precisely estimated. As we explain later, this null finding is neither surprising nor in conflict with the prior evidence when interpreted in the context of local proportional representation (PR) elections.

Since the seminal paper on RDD by Hahn, Todd, and van der Klaauw (2001), non-parametric local linear regression has been used widely in applied work to approximate the regression function near the cutoff. A key decision in implementing local methods is the choice of a bandwidth, which defines how close to the cutoff the estimation is implemented; various methods have been proposed for selecting it (e.g., Ludwig and Miller (2007), Imbens and Kalyanaraman (2012), and Calonico, Cattaneo, and Titiunik (2014a); see also Calonico, Cattaneo, and Farrell (2016a)). For example, a mean-squared-error (MSE) optimal bandwidth trades off the bias due to not getting the functional form completely right for wider bandwidths with the increased variance of the estimate for narrower bandwidths. We find that when RDD is applied to our elections data and implemented in the conventional fashion using local-polynomial inference with MSE-optimal bandwidths, the estimates indicate a statistically significant positive personal incumbency advantage. This finding means that the conventional implementation, which still appears to be the preferred approach by many practitioners, can lead to misleading results.

The disparity between the experimental and the RDD estimates suggests that the implementation of RDD using local-polynomial inference with MSE-optimal bandwidths is deficient.⁴ Local methods may produce biased estimates if the parametric specification is not a good approximation of the true regression function within the bandwidth (e.g., Imbens and Lemieux (2008)).⁵ If the bias is relatively large, the MSE-optimal bandwidth does not provide a reliable basis for inference, as it then produces confidence intervals that have incorrect asymptotic coverage (Calonico, Cattaneo, and Titiunik (2014a)).

We find that when an ad hoc undersmoothing procedure of using smaller (than MSE-optimal) bandwidths is used to reduce the bias (see, e.g., Imbens and Lee (2008) and Calonico, Cattaneo, and Farrell (2016a)), the null hypothesis of no personal incumbency advantage is no longer rejected. However, we cannot determine whether this is due to better size properties or wider confidence intervals (inefficiency). More importantly, we show that the bias-correction and robust inference procedure of Calonico, Cattaneo, and Titiunik (2014a) brings the RDD estimate(s) in line with the experimental

7, 2014), and in Norway as well as in Canada and the United Kingdom (http://en.wikipedia.org/wiki/Coin_flipping#Politics). We acknowledge that in some of these elections ties are probably too rare for a meaningful statistical analysis, but this nevertheless hints at the possibility of carrying out similar comparisons in other countries. At least in countries where a similar open list system is used at the local level, there should be enough ties to replicate our analysis. For example, Chile and Colombia might be such countries.

⁴Another potential reason why the experimental estimate and the estimate that our standard implementation of a close election RDD generates do not match is that the conditional expectation of the potential outcome is not continuous at the cutoff. We find no signs of this key RDD assumption being violated using covariate balance checks.

⁵In our case, curvature is clearly visible within the bandwidth optimized for the local linear specification.

estimate, provided that one does not allow for too large a bandwidth for bias estimation. This finding is important for applied RDD analysis, as this implementation of RDD corrects for the bias in the confidence intervals and results in narrower confidence intervals (implying more power than the ad hoc undersmoothing procedures) that have faster vanishing coverage error rates (see also Calonico, Cattaneo, and Farrell (2016a)). Given that we build on a real-world experiment, we provide an independent verification of the empirical performance of the Calonico, Cattaneo, and Titiunik (2014a) procedure: We find that the procedure is less sensitive to the choice of the bandwidth (than ad hoc undersmoothing) and works especially well when the bandwidth used for bias estimation (“bias bandwidth”) and the bandwidth used to estimate the regression discontinuity effect (“RD effect bandwidth”) are set equal. These findings support the results of Monte Carlo simulations and formal analyses reported in Calonico, Cattaneo, and Titiunik (2014a) and Calonico, Cattaneo, and Farrell (2016a). Our evidence complements these Monte Carlo results, as the experimental estimate provides an alternative benchmark against which RDD can be compared. Unlike the benchmark provided by the Monte Carlo, our approach (like LaLonde 1986) does not force the econometrician to assume that the true data generating process is known.

We also find, in line with the prior work, that using richer local polynomial specifications for a given bandwidth optimized for the linear specification can eliminate the bias. However, when higher order local polynomials are used and the bandwidths are accordingly optimized, the bandwidths tend to become too large and the bias typically remains. This implies that MSE-optimal bandwidths may be problematic more generally. Consistent with this, the recent work of Calonico, Cattaneo, and Farrell (2016a) suggests that a particular bandwidth adjustment (“shrinkage”) is called for to achieve better coverage error rates and more power when MSE-optimal bandwidths are used.

Echoing Calonico, Cattaneo, and Titiunik (2014a), we provide a word of caution to practitioners, since the local (linear) regression with an MSE-optimal bandwidth, which is often used in applied work, appears to lead to an incorrect conclusion. Our results show that careful implementation of the bias-correction and robust inference procedure of Calonico, Cattaneo, and Titiunik (2014a) can meet the replication standard in the context of close elections.

Previous work has compiled a good body of evidence about how valid the RDD identification assumptions are in various contexts, including elections. However, this paper is, to our knowledge, the first to provide direct evidence of the remaining fundamental question of how well the various RDD estimation techniques perform that is separate from the questions of identification. That is, how well these various approaches work when the identification assumptions are met? Our results demonstrate that the inferences in RDD can be sensitive to the details of the implementation approach even when the sample size is relatively large.

Our empirical analysis also bears on four other strands of the literature. First, there is an emerging literature on within-study comparisons of RDDs to experiments (Black, Galdo, and Smith (2007), Cook and Wong (2008), Cook, Shadish, and Wong (2008), Green et al. (2009), Shadish et al. (2011), and Wing and Cook (2013)) that explores how the performance of RDD depends on the context in which it is used. A key limitation of all these

studies is that the experimental treatment effects are different from the one that RDD targets. Moreover, they do not use the most recent RDD implementations.⁶ Thus, while insightful, it is unclear how relevant these prior papers are for the currently ongoing RDD development efforts. Second, it has been argued that in close elections, the conditions for covariate balance (and local randomization) around the cutoff do not necessarily hold, especially in post-World War II U.S. House elections (Snyder (2005), Caughey and Sekhon (2011), and Grimmer et al. (2011)). Eggers et al. (2015) convincingly challenge this conclusion (see also Erikson and Rader (2017)).⁷ We contribute to this ongoing debate by showing whether and when the close election RDD is capable of replicating the experimental estimate. Third, we provide evidence that the local randomization approach advocated by Cattaneo, Frandsen, and Titiunik (2015) is also able to replicate the experimental estimate. Finally, our findings add to the cumulating evidence on limited personal incumbency advantage in proportional representation (PR) systems (see, e.g., Lundqvist (2013), Redmond and Regan (2015), Golden and Picci (2015), Dahlgard (2016), and Kotakorpi, Poutvaara, and Terviö (2017)).

The rest of this paper is organized as follows. In Section 2, we describe the institutional environment and our data. The experimental and non-experimental results are reported and compared in Section 3. We discuss the validity and robustness of our findings in Section 4. Section 5 concludes. A large number of additional analyses are reported in a supplemental appendix available in a supplementary file on the journal website, http://qeconomics.org/supp/864/code_and_data.zip.

2. INSTITUTIONAL CONTEXT AND DATA

2.1 *Institutional environment*

Finland has a two-tier system of government that consists of a central government and a large number of municipalities at the local level.⁸ Finnish municipalities have extensive tasks and considerable fiscal autonomy. In addition to the usual local public goods and services, municipalities are responsible for providing most of social and health care services and primary and secondary education. Municipalities are therefore of considerable importance to the whole economy.⁹

Municipalities are governed by municipality councils. The council is by far the most important political actor in municipal decision making. For example, mayors are public

⁶The current view of this literature is that RDD is able to reproduce—or at least to approximate—experimental results in most, but not in all, settings (see Cook, Shadish, and Wong (2008) and Shadish et al. (2011)). There are also a number of unpublished working papers on this topic, but they suffer from the same limitations as the published ones.

⁷The criticism on the close election RDD builds on the argument that outright fraud, legal and political manipulation, and/or sorting of higher quality or better positioned candidates may naturally characterize close elections. However, Eggers et al. (2015) show that post-World War II U.S. House elections are a special case and that there is no imbalance in any of the other elections that their data set on 40,000 close political races cover.

⁸In 1996, Finland had 436 municipalities and in 2012, 304.

⁹Municipalities employ around 20% of the total workforce. The most important revenue sources of the Finnish municipalities are local income taxes, operating revenues, such as fees, and funding from the central government.

officials chosen by the councils and can exercise only partial executive power. Moreover, municipal boards (i.e., cabinets) have a preparatory role only. The party presentation in the boards follows the same proportional political distribution as the presentation in the council.

Municipal elections are held simultaneously in all municipalities. All municipalities have one electoral district. The council size is determined by a step function based on the municipal population. The median council size is 27. The elections in our data were held on the fourth Sunday of October in 1996, 2000, 2004, 2008, and 2012. The four-year council term starts at the beginning of the following year. The seat allocation is based on PR, using the open-list D'Hondt election rule. There are three (1996–2008 elections) or four (2012 elections) major parties, which dominate the political landscape of both the municipal and the national elections, as well as four other parties that are active both locally and nationally. Moreover, some purely local independent political groups exist. In the elections, each voter casts a single vote to a single candidate. One cannot vote for a party without specifying a candidate. In this setting, voters (as opposed to parties) decide which candidates are eventually elected from a given list, because the number of votes that a candidate gets determines the candidate's rank on her party's list.

The total number of votes over the candidates of a given party list determines the votes for each party. The parties' votes determine how many seats each party gets. The procedure is as follows. First, a comparison index, which equals the total number of votes cast to a party list divided by the order (number) of a candidate on the list, is calculated for all the candidates of all the parties. The candidates are then ranked according to the index and all those who rank higher than $(S + 1)$ th (S being the number of council seats) get a seat.

An important feature of this election system is that in many cases, there is an exact tie in the number of votes at the margin where the last available seat for a given party list is allocated. This means that within a party, the rank of two or more candidates has to be randomly decided. For example, it is possible that a party gets k seats in the council and that the k th and $(k + 1)$ th ranked candidates of the party receive exactly the same number of votes. For them, the comparison index is the same. The applicable Finnish law dictates that in this case, the winner of the marginal (k th) seat has to be decided using a randomization device. Typically, the seat is literally allocated by drawing a ticket (name) from a hat. The procedure appears to be very simple: One of the (typically female) members of the municipal election committee wears a blindfold and draws the ticket in the presence of the entire committee.¹⁰ While we have not run an experiment or implemented a randomized controlled trial, we can use the outcomes from these lotteries to generate an experimental treatment effect estimate for the effect of incumbency status on electoral support.

It is also possible that two (or more) candidates from *different parties* face a tie for a marginal seat. However, within-party ties are much more common in practice. Therefore, we do not analyze ties between candidates from different parties. Besides resulting in a larger sample in which the candidates had a tie, there are three additional reasons

¹⁰See, for example, an article in one of the major Finnish tabloids, *Iltasanomat*, on April 12, 2011.

to focus on the within-party ties. First, using the within-party ties allows for a simpler implementation of RDD, as we do not have to worry about discontinuities and possible party-level incumbency effects that are related to party lines.¹¹ Second, focusing on the within-party dimension allows a cleaner identification of the personal incumbency effect, net of the party incumbency effect. Third, the use of within-party ties increases the comparability of our RDD analysis, which uses multiparty PR elections data, with the prior studies that use data from two-party (majoritarian) systems. This is so as within a party list, the Finnish elections follow the N -past-the-post rule. In both cases, personal votes determine who gets elected.

2.2 Data

Our data originate from several sources. The first source is election data provided by the Ministry of Justice. These data consist of candidate-level information on the candidates' age, gender, party affiliation, the number of votes they received, their election outcomes (elected status), and the possible incumbency status. The election data were linked to data from Kuntien eläkevakuutus (KEVA; formerly known as the Local Government Pensions Institution) to identify municipal workers, and to Statistics Finland's data on the candidates' education, occupation, and socioeconomic status. We further added income data from the Finnish tax authority. Finally, we matched the candidate-level data with Statistics Finland's data on municipal characteristics.¹²

We have data on 198,121 candidates from elections held in years 1996, 2000, 2004, 2008, and 2012.¹³ Summary statistics (reported in Appendix A) show that the elected candidates differ substantially from those who are not elected: They have higher income and more often a university degree, and are less often unemployed. The difference is particularly striking when we look at the incumbency status: 58% of the elected candidates were incumbents, whereas only 6% of those who were not elected were incumbents.

3. MAIN RESULTS

3.1 *Experimental estimates*

In this section, we estimate the magnitude of the personal incumbency advantage using the data from the random election outcomes. We define this added electoral support as the treatment effect of getting elected today on the probability of getting elected in the next election. We measure the event of getting elected today by a binary indicator, Y_{it} , which takes value of 1 if candidate i was elected in election year t and is 0 otherwise. Our

¹¹See Folke (2014) for the complications that multiparty systems generate and Snyder, Folke, and Hirano (2015) on issues with partisan imbalance in RDD studies.

¹²The candidate-level demographic and occupation data usually refer to the election year, with the exception that occupation data from 1995 (2011) are matched to 1996 (2012) elections data.

¹³Two further observations on the data are in order. First, to be careful, we omit all data (about 150 candidates) from one election year (2004) in one municipality, because of a mistake in the elected status of one candidate. The mistake is apparently due to one elected candidate being disqualified later. Second, the data on the candidates running in 2012 are only used to calculate the outcome variables.

main outcome is a binary variable, $Y_{i,t+1}$, which equals 1 if candidate i is elected in the next election year $t + 1$ and is 0 otherwise.

In elections between 1996 and 2008, 1351 candidates had a tie within their party lists for the last seat(s), that is, at the margin that determines whether or not the candidates get a seat.¹⁴ In these cases, a lottery was used to determine who got elected. This implies that Y_{it} was randomly assigned in our *lottery sample*, that is, among the candidates who had a tie.

3.1.1 Covariate balance tests for the lottery sample Was the randomization required by the law conducted correctly and fairly? To address this question, we study whether candidates' characteristics balance between the treatment (randomly elected) and the control group (randomly not elected) within the lottery sample. The results are reported in Table 1. The differences are statistically insignificant and small in magnitude. These findings support the view that Y_{it} is randomly determined in the lottery sample.¹⁵

3.1.2 Experimental estimate for the personal incumbency effect Is there a personal incumbency effect? Before we can answer this question, we have to point out that a subsequent electoral outcome is observed for 820 out of the 1351 candidates who participated in the lottery between 1996 and 2008, because they reran in a subsequent election. We do not know what happened to those who decided not to rerun. This attrition is a possible source of concern, because the decision not to rerun may mirror, for example, the candidates' expected performance. If it does, analyses based on the selected sample, from which those who did not rerun are excluded, would not provide us with the correct treatment effect. Rerunning is an (endogenous) outcome variable and we therefore cannot condition on it unless the treatment has no effect on the likelihood of rerunning. Relying on such an assumption would be neither harmless nor conservative.¹⁶ Our baseline results therefore refer to the entire lottery sample. This means that we code our main outcome variable so that it is equal to 1 if the candidate is elected in the next election and is set to 0 if the candidate is not elected or does not rerun.

The fraction of candidates who get elected in election year $t+1$ conditional on not winning the lottery in election year t is 0.325, whereas the same fraction conditional on winning the lottery is 0.329. The difference between the two fractions provides us with a first experimental estimate of personal incumbency advantage. It is small, ≈ 0.004 . Because Y_{it} is randomly assigned in the lottery sample, the difference estimates the average treatment effect (ATE). Note that due to the way the lottery sample is constructed,

¹⁴In addition, there were 202 ties in 2012. We do not include them in the lottery sample, because we do not have data on the subsequent election outcomes for these candidates. When we include these ties in the balancing tests, the results do not change. Notice also that a tie may involve more than two candidates and more than one seat. For example, three candidates can tie for two seats.

¹⁵The candidates' party affiliations and municipal characteristics should be balanced by design, because we analyze lotteries within the party lists. The corresponding balancing tests (reported in Appendix B) confirm this.

¹⁶In the party level analysis of Klačnja and Titunik (2016), the dependent variable is a binary variable equal to 1 if the party wins the election at $t + 1$ and equal to 0 if the party either runs and loses at $t + 1$ or does not run at $t + 1$. Similar to ours, their main analysis includes all observations (i.e., does not condition on whether a party reruns). In an appendix, Klačnja and Titunik (2016) also report an analysis conditioning on running again.

TABLE 1. Covariate balance tests for the lottery sample.

Variable	Elected ($N = 671$)			Not Elected ($N = 680$)			Difference	p -Value	p -Value (Clustered)
	N	Mean	Std. Dev.	N	Mean	Std. Dev.			
Vote share	671	1.54	0.69	680	1.53	0.67	0.00	0.93	0.97
Number of votes	671	41	39	680	41	38	0	0.83	0.93
Female	671	0.39	0.49	680	0.38	0.49	0.01	0.80	0.80
Age	671	45.42	11.87	680	45.69	11.54	-0.27	0.67	0.67
Incumbent	671	0.29	0.45	680	0.31	0.46	-0.02	0.34	0.35
Municipal employee	671	0.24	0.43	680	0.25	0.44	-0.01	0.62	0.62
Wage income	478	22,521	14,928	476	22,256	13,729	265	0.78	0.82
Capital income	478	2929	18,612	476	3234	12,085	-305	0.76	0.81
High professional	671	0.18	0.38	680	0.18	0.38	0.00	0.97	0.97
Entrepreneur	671	0.24	0.43	680	0.24	0.43	0.00	0.84	0.87
Student	671	0.02	0.15	680	0.03	0.16	0.00	0.76	0.76
Unemployed	671	0.06	0.24	680	0.05	0.22	0.01	0.37	0.37
University degree	537	0.13	0.34	545	0.13	0.34	0.00	0.86	0.86

Note: The difference in means has been tested using a t -test with and without clustering at the municipality level. The sample includes only candidates running in 1996–2008 elections. For 1996, income data are available only for candidates who run also in 2000, 2004, or 2008 elections. Wage and capital income are annual and are expressed in nominal euros.

TABLE 2. Experimental estimates of the personal incumbency advantage.

	(1)	(2)	(3)	(4)
Elected	0.004	0.001	-0.010	-0.010
95% confidence interval (robust)	[-0.046, 0.054]	[-0.049, 0.051]	[-0.064, 0.040]	[-0.060, 0.040]
95% confidence interval (clustered)	[-0.044, 0.053]	[-0.048, 0.050]	[-0.067, 0.047]	[-0.075, 0.055]
N	1351	1351	1351	1351
R^2	0.00	0.03	0.28	0.44
Controls	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	No
Municipality-year fixed effects	No	No	No	Yes

Note: Only actual lotteries are included in the regressions. Set of controls includes age, gender, party affiliation, socioeconomic status and incumbency status of a candidate, and total number of votes. Some specifications include also municipality or municipality-year fixed effects. Confidence intervals based on clustered standard errors account for clustering at the municipality level. The unit of observation is a candidate i at year t .

this ATE is estimated *precisely at the cutoff point* of political support that determines whether or not a candidate gets elected. It is therefore an ideal benchmark for the non-experimental RDD estimate, because the sharp RDD targets exactly the same treatment effect.

To perform inference (and to provide a set of complementary experimental estimates), we regress $Y_{i,t+1}$ on Y_{it} using ordinary least squares (OLS) and the sample of candidates who faced within-party ties. Table 2 reports the point estimates and 95% confidence intervals that are robust to heteroscedasticity and, separately, that allow for clustering at the level of municipalities. In the leftmost column, $Y_{i,t+1}$ is regressed on Y_{it} and a constant using OLS. The coefficient of Y_{it} is 0.004, as expected. The estimate is statistically insignificant: Both 95% confidence intervals include 0. The estimate is insignificant also if a conventional (nonrobust, nonclustered) t -test is used: The p -value of the standard t -test is 0.87. In the remaining columns, we report the OLS results from a set of specifications that include control variables and fixed effects. Three main findings emerge. First, there is no evidence of a personal incumbency advantage. The estimated effect is close to 0 across the columns and the 95% confidence intervals always include 0. Second, the coefficient of Y_{it} is relatively stable across the columns and is thus not correlated with the added controls or fixed effects. This further supports the view that Y_{it} is random. Third, the confidence intervals are fairly narrow. For example in specification (1), effects larger than 5.3 percentage points are outside the upper bound of the clustered confidence interval. We can thus at least rule out many of the (much) larger effects typically reported in the incumbency advantage literature on majoritarian elections.

We have considered the robustness of the experimental estimate(s) in various ways. First, 0.9% of the candidates run in another municipality in the next election. For Table 2, they were coded as rerunning. The results (not reported) are robust to coding them as not re-running. Second, we have considered the vote share in the next election as an alternative outcome. While more problematic, we follow the same practice with this alternative outcome as above and set it to 0 if the candidate did not rerun in the next election. The results (reported in Appendix B) show that Y_{it} has no impact on the vote share.

Third, we have studied small and large elections separately (see Appendix B). We still find no evidence of a personal incumbency advantage. Finally, we get an experimental estimate close to 0 (for both the elected next election and vote share next election outcomes) if we use a trimmed lottery sample that only includes the rerunners (reported in Appendix B).

We have also checked that when the event of rerunning in the next election is used as the dependent variable, the experimental estimate is small and statistically not significant (see Appendix B). The past winners are therefore not more (or less) likely to rerun, giving credence to the view that the treatment effect on which we focus is a valid estimate of the incumbency effect.

3.1.3 Discussion of the experimental estimate The personal incumbency advantage refers to the added electoral support that an incumbent politician of a given party enjoys when she runs for re-election, relative to her being a non-incumbent candidate from the same party and constituency.¹⁷ Such an advantage could stem from various sources, such as from having been able to serve the constituency well, having enjoyed greater public visibility while holding the office, improved candidate quality (through learning while in power), reduced competitor quality (due to a “scare-off” effect; see Cox and Katz (1996) and Erikson and Titiunik (2015)), and the desire of voters to disproportionately support politicians with past electoral success (“winners”). The earlier (mostly U.S.) evidence suggests that the existence of an incumbent advantage in two-party systems is largely beyond question (see, e.g., Erikson and Titiunik (2015) and references therein). It is clear that the size of the advantage may nevertheless vary and be context-specific; see, for example, Desposato and Petrocik (2003), Grimmer et al. (2011), Uppal (2009), and Klašnja and Titiunik (2016), who find evidence of a party-level disadvantage in systems characterized by weak parties.

In our view, the null finding of no personal incumbency advantage is neither surprising nor in conflict with the prior evidence, for two reasons. First, we are looking at personal incumbency advantage in the context of small local PR elections. It is possible that in this context, the randomized political victories take place at a relatively unimportant margin. For example, such a political win does not, per se, typically lead to a visible position in media or a prominent position in the wider political landscape. Perhaps being the last elected candidate of a party in the Finnish municipal elections conveys limited opportunities to serve one’s constituency or to improve one’s quality as a candidate through learning-by-doing.¹⁸ What is more, it is certainly plausible that getting the last seat by a lottery or by only a very small margin does not work to scare off good competitors in the subsequent elections. Such a political victory provides the voters with a

¹⁷The party incumbency advantage, in turn, measures the electoral gain that a candidate enjoys when she is from the incumbent party, independently of whether she is an incumbent politician or not (Gelman and King (1990) and Erikson and Titiunik (2015)). Following Lee (2008), most of the earlier RDD analyses refer to the party advantage (e.g., Broockman (2009), Caughey and Sekhon 2011, and Trounstine 2011; see also Fowler and Hall (2014)).

¹⁸Similarly, being the first non-elected candidate of a party may convey some opportunities to participate in the municipal decision making, for example, by serving as a deputy councilor or as a member of municipal committees.

limited opportunity to picture and support the candidate as a political winner. It is thus not surprising if there is no personal incumbency advantage at the margin that we study.

Second, it is important to recall that most of the recent RDD evidence on the positive and large incumbency effects mirrors both a party and a personal effect.¹⁹ In contrast, the random election outcomes in our data allow recovering a treatment effect estimate for the personal incumbency advantage that specifically excludes the party effect, because it is estimated from within-party variation in the incumbency status. If the party effect is positive, the effects we find are likely to be lower than what has been reported in prior work. Moreover, the existing studies that look at a personal incumbency advantage in the PR systems of developed countries typically find only modest or no incumbency effects (Lundqvist (2013), Golden and Picci (2015), Dahlgard (2016), and Kotakorpi, Poutvaara, and Terviö (2017)).

3.2 *Non-experimental estimates*

3.2.1 *Implementing RDD for PR elections* Our forcing variable is constructed as follows. We measure closeness *within a party list* so as to focus on the same cutoff where the lotteries take place, and to abstract from multiparty issues in constructing the forcing variable and potential party effects in PR systems (see Folke (2014)). To this end, we calculate for each ordered party list the pivotal number of votes as the average of the number of votes among the first non-elected candidate(s) and the number of votes among the last elected candidate(s). A candidate's distance from getting elected is then the number of votes she received minus the pivotal number of votes for her list (party). We normalize this distance by dividing it by the number of votes that the party list got in total and then multiply it by 100.²⁰ This normalized distance is our forcing variable v_{it} .²¹

Four observations about our forcing variable are in order. First, it measures closeness within a party list in vote shares. It is thus in line with the existing measures for majoritarian systems. As usual, all candidates with $v_{it} > 0$ get elected, whereas those with $v_{it} < 0$ are not elected. All those candidates for whom $v_{it} = 0$ face a tie and get a seat if they win in the lottery. Second, the histogram of the forcing variable close to the cutoff (reported in Appendix C) shows that there are observations close to the cutoff and thus that some, but not extensive, extrapolation is being done in the estimation of the RDD treatment effect. Third, the assumption of having a continuous forcing variable is not at odds with our forcing variable. For example, among the 100 closest observations to the cutoff, 92 observations obtain a unique value of v_{it} and there are 4 pairs for which the value is the same within each pair. Finally, our normalized forcing variable and the

¹⁹These two effects cannot typically be distinguished from each other unless parametric assumptions are made (Erikson and Titiunik (2015)).

²⁰This definition of the forcing variable means that all those party lists from which no candidates or all candidates got elected are dropped from the analysis. In total, this means omitting about 6000 candidate-election observations. This corresponds to roughly 3% of the observations in the elections organized between 1996 and 2012.

²¹Dahlgard (2016), Golden and Picci (2015), Lundqvist (2013), and Kotakorpi, Poutvaara, and Terviö (2017) study quasi-randomization that takes place within parties in a PR system using an approach similar to ours.

(potential alternative) forcing variable based on the absolute number of votes operate on a very different scale, but they are correlated (their pairwise correlation in our data is 0.34, p -value < 0.001 ; see also Appendix C).²² Moreover, as we discuss later in connection with robustness tests, our RDD results are robust to using alternative definitions of the forcing variable.

A special feature of a PR election system is that it is much harder than in a two-party majoritarian system for a candidate or a party to accurately predict the precise location of the cutoff that determines who gets elected from a given party list. The reason for this is that the number of seats allocated to the party also depends on the election outcome of the other parties. This makes it more likely that the forcing variable cannot be perfectly manipulated.

The function of the forcing variable is estimated separately for both sides of the cutoff. The choice of the bandwidth determines the subsample near the cutoff to which the function of the forcing variable is fitted and from which the treatment effect is effectively estimated (Imbens and Lemieux (2008), Lee (2008), and Lee and Lemieux (2010)). For our baseline RDD, we use a triangular kernel and the widely used implementations of the (MSE-optimal) bandwidth selection of Imbens and Kalyanaraman (2012; IK hereafter) and Calonico, Cattaneo, and Titiunik (2014a; CCT hereafter).²³ We report results from a sharp RDD for the subsample of candidates that excludes the randomized candidates, because a typical close election RDD would not have such lotteries in the data.

3.2.2 RDD estimations: Graphical analysis We start by displaying the relationship between the forcing variable and the outcome variable close to the cutoff in Figure 1.²⁴ The graph suggests that there is substantial curvature in this relation. In Panel A, the width of the x -axis is 1 IK bandwidth of the local linear specification on both sides of the cutoff. The fits are those of local linear (on the left), quadratic (in the middle), and cubic (on the right) regressions. The figure on the left clearly shows that there is curvature in the data near the cutoff, making the linear approximation inaccurate. This finding is not due to using the linear probability model, as Logit and Probit models generate similar

²²In large elections, it is more likely that small vote share differences are observed (rather than small differences in the number of votes). The opposite holds for small elections.

²³Two further points are worth mentioning here. First, the IK and CCT bandwidths are two different implementations of the estimation of the MSE-optimal theoretical bandwidth choice (i.e., the one that optimizes the asymptotic mean-squared-error expansion). The older (2014) version of the Stata software package `rdrobust` (developed by Calonico, Cattaneo, and Titiunik (2014a, 2014b)) offered the possibility of using these two bandwidth selectors. In the upgraded version of the package, the IK and CCT bandwidth selectors have been deprecated. The upgraded version now uses a third implementation of the estimation of the MSE-optimal theoretical bandwidth choice (see Appendix E). Second, we have also calculated the bandwidths proposed by Fan and Gijbels (1996) and Ludwig and Miller (2007). As those were always broader than both the IK and CCT bandwidths and are currently less often used in practice, we do not report them.

²⁴The figure has been produced by the `rdplot` command for Stata that approximates the underlying unknown regression functions without imposing smoothness (Calonico, Cattaneo, and Titiunik (2015)). The key contribution in Calonico, Cattaneo, and Titiunik (2015) is to provide a data-driven approach for choosing the bin widths that allows bin sizes to vary, instead of using ad hoc bins of equal sizes. In Appendix C, we provide an alternative version of Figure 1 with a richer illustration of the raw data.

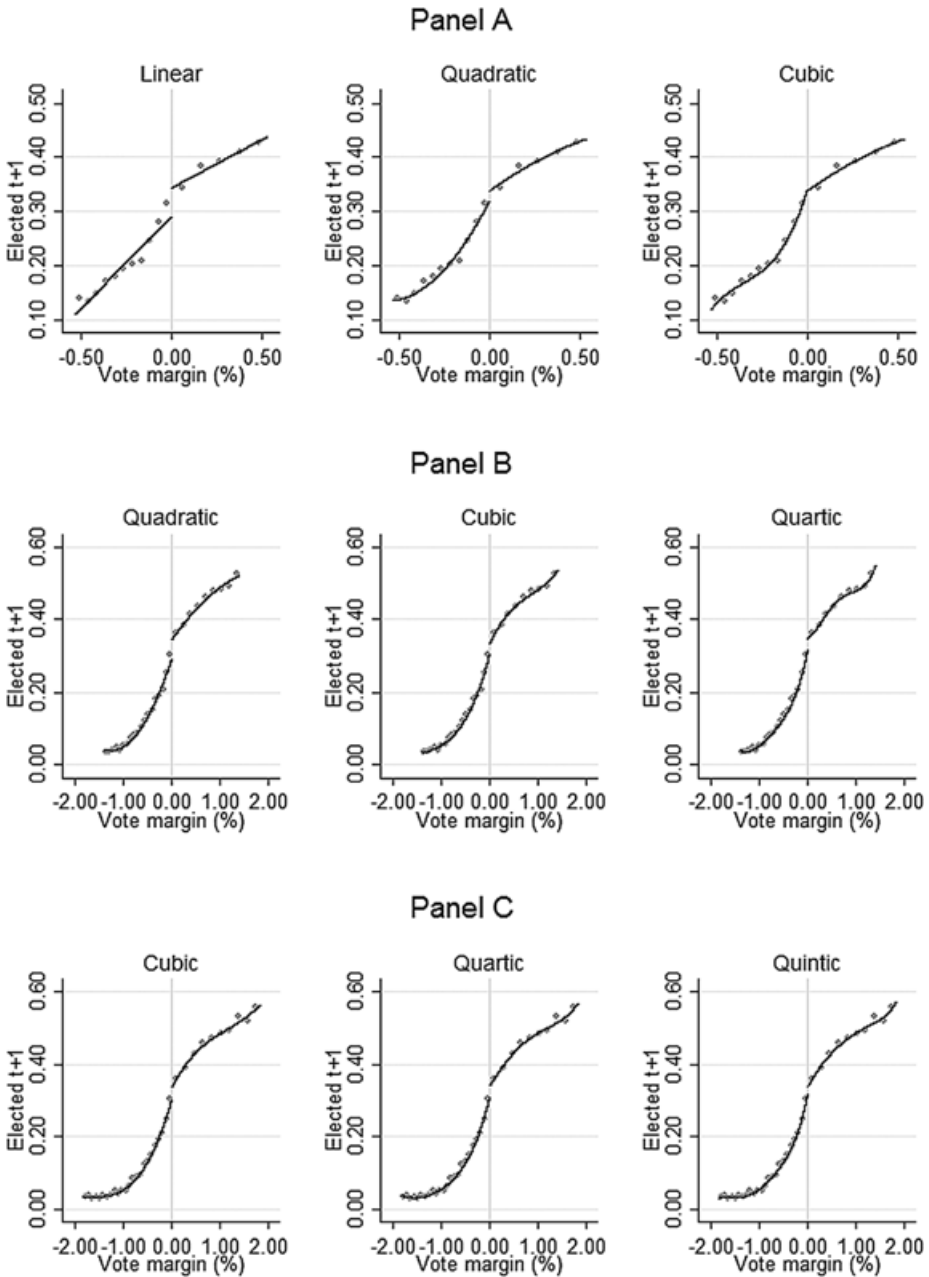


FIGURE 1. Curvature between the forcing variable and the outcome. The figure shows local polynomial fits based on a triangular kernel and the IK bandwidth. The IK bandwidth was optimized for the linear specification in Panel A, for the quadratic specification in Panel B, and for the cubic specification in Panel C. On left side, the graphs display the fits that are based on the same p (order of local polynomial specification) as that for which optimal bandwidths are calculated. In the center graph, the fit uses a $p + 1$ specification and on the right side, the graphs are based on a $p + 2$ specification. Gray dots mark binned averages where the bins are chosen using the integrated MSE-optimal evenly spaced approach of Calonico, Cattaneo, and Titiunik (2015).

insights (not reported). The quadratic local polynomial in the middle seems to capture the curvature quite well. This finding suggests that a polynomial specification of order 2 might be flexible enough for the bandwidth that has been optimized for a polynomial of order 1.

The same observation can be made from Panels B and C of Figure 1, where the bandwidths are optimal for the quadratic (Panel B) and cubic (Panel C) specifications. As in Panel A, the graphs on the left hand side of these panels display the fits that are based on the same order of the local polynomial specification, p , for which the optimal bandwidth is calculated. In the middle graph, the fit uses a $p + 1$ local polynomial, but the bandwidth is the same as on the left hand side. In the graphs on the right hand side, the displayed fits are based on a $p + 2$ local polynomial. The approximation is better especially near the cutoff when the richer $p + 1$ polynomial is used. Moreover, the experimental estimate indicates that there should not be a jump at the cutoff. The graphs on the left are therefore consistent with a poor local approximation, because there a jump can be detected. The jumps are nearly invisible or completely nonexistent in the graphs displayed in the middle ($p + 1$) or on the right ($p + 2$).²⁵

3.2.3 RDD estimations: Baseline results Table 3 reports our baseline RDD estimation results. In each panel of the table, we report the conventional RDD point estimates and the 95% confidence intervals that are robust to heteroscedasticity and, separately, that allow for clustering at the level of municipalities.²⁶ The panels differ in how the bandwidths and local polynomials are used.

In Panel A of Table 3, the bandwidth is selected optimally for the local linear specification using either the IK or the CCT implementation of the bandwidth selection. For these bandwidth choices, the panel reports the local linear (specifications (1) and (2)), quadratic (specifications (3) and (4)), and cubic (specifications (5) and (6)) RDD estimates of the personal incumbency advantage. As specifications (1) and (2) show, both local linear RDD specifications with bandwidths that are optimally chosen for the linear specification indicate a positive and statistically significant incumbency advantage. The local linear RDD with optimal bandwidth is thus not able to replicate the experimental estimate. This is likely to happen when the regression function has curvature within the optimal bandwidth that the linear approximation cannot capture. The next specifications (specifications (3)–(6)) in the panel show that the curvature of the regression function indeed matters. Using the richer quadratic and cubic local polynomials aligns the RDD estimates with the experimental results for the bandwidths that are MSE-optimal,

²⁵We checked that a polynomial specification $p + 1$ is flexible enough for bandwidth optimized for p from $p = 0$ to $p = 5$ in our case. We have also checked that these findings are not specific to the way we define the forcing variable. The same patterns can be observed also if we use the absolute number of votes as the forcing variable (reported in Appendix C).

²⁶We report the confidence intervals that are robust to heteroscedasticity only (i.e., that do not allow for clustering), because the bandwidth selection techniques are not optimized for clustered inference. On the other hand, clustering is common among applied researchers. We therefore also report cluster-robust confidence intervals (but acknowledge that the choice of the clustering unit is hard to justify). See Bartalotti and Brummet (2017) for a recent analysis of cluster-based inference in the context of RDD.

TABLE 3. Local polynomial RDD estimates: the outcome is elected next election.

Panel A: Bandwidth Optimized for Local Linear Specification						
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear		Quadratic		Cubic	
Elected	0.039	0.052	0.008	0.022	-0.021	-0.004
95% confidence interval (robust)	[0.009, 0.070]	[0.027, 0.078]	[-0.038, 0.055]	[-0.018, 0.062]	[-0.087, 0.045]	[-0.057, 0.049]
95% confidence interval (clustered)	[0.011, 0.068]	[0.028, 0.077]	[-0.037, 0.053]	[-0.015, 0.059]	[-0.086, 0.044]	[-0.056, 0.048]
<i>N</i>	19,648	27,218	19,648	27,218	19,648	27,218
<i>R</i> ²	0.03	0.05	0.03	0.05	0.03	0.05
Bandwidth	0.54	0.74	0.54	0.74	0.54	0.74
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT
Panel B: 0.5 * Bandwidth Optimized for Local Linear Specification						
	(7)	(8)	(9)	(10)	(11)	(12)
	Linear		Quadratic		Cubic	
Elected	0.008	0.025	-0.022	-0.015	-0.019	-0.025
95% confidence interval (robust)	[-0.038, 0.053]	[-0.013, 0.062]	[-0.093, 0.049]	[-0.074, 0.045]	[-0.125, 0.088]	[-0.107, 0.057]
95% confidence interval (clustered)	[-0.035, 0.050]	[-0.011, 0.060]	[-0.094, 0.050]	[-0.072, 0.043]	[-0.128, 0.091]	[-0.110, 0.060]
<i>N</i>	9934	13,586	9934	13,586	9934	13,586
<i>R</i> ²	0.01	0.02	0.02	0.02	0.02	0.02
Bandwidth	0.27	0.37	0.27	0.37	0.27	0.37
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT

(Continues)

TABLE 3. *Continued.*

Panel C: Bandwidths Optimized for Each Specification Separately						
	(13)	(14)	(15)	(16)	(17)	(18)
	Linear		Quadratic		Cubic	
Elected	0.039	0.052	0.039	0.057	0.030	0.055
95% confidence interval (robust)	[0.009, 0.070]	[0.027, 0.078]	[0.013, 0.066]	[0.036, 0.078]	[-0.002, 0.062]	[0.035, 0.076]
95% confidence interval (clustered)	[0.011, 0.068]	[0.028, 0.077]	[0.013, 0.065]	[0.036, 0.078]	[-0.000, 0.060]	[0.035, 0.076]
N	19,648	27,218	54,436	78,469	70,576	112,398
R^2	0.03	0.05	0.11	0.16	0.15	0.23
Bandwidth	0.54	0.74	1.41	2.09	1.84	3.98
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT

Note: The table shows estimated incumbency advantage using local polynomial regressions within various bandwidths. All estimations use a triangular kernel. Confidence intervals based on clustered standard errors account for clustering at the municipality level. The unit of observation is a candidate i at year t . The IK and CCT bandwidths are two different implementations of the estimation of the MSE-optimal theoretical bandwidth choice.

as determined by IK and CCT implementations of the MSE-optimal bandwidth *for the linear specification*.²⁷

In Panel B of the table, we report the results using bandwidths that are half the optimal bandwidth of the local linear specification. This undersmoothing ought to reduce the (asymptotic) bias, which it indeed appears to do. When the local linear polynomial specification and bandwidths half the size of the IK or CCT bandwidths are used, the point estimates decrease in size and the results are in line with the experimental benchmark (specifications (7) and (8)). The null hypothesis of no personal incumbency effect also cannot be rejected when the quadratic and cubic polynomials are used (specifications (9)–(11)).

Finally, in Panel C, we report the results for the quadratic and cubic specifications, with IK and CCT implementations of the MSE-optimal bandwidths that have been re-optimized for these more flexible specifications. As the panel shows, we find, bar one exception, positive and statistically significant effects.

These findings are consistent with the view that when the MSE-optimal bandwidths are used in the local polynomial regression, there is a risk of overrejection because the distributional approximation of the estimator is poor (Calonico, Cattaneo, and Titiunik (2014a)). What also is in line with the recent econometric work is that holding the order of the polynomial constant, smaller bandwidths align our RDD results with the experimental benchmark (see Calonico, Cattaneo, and Titiunik (2014a) for a discussion of undersmoothing).²⁸ Moreover, we find that holding the bandwidth constant, richer polynomials align our RDD results with the experimental benchmark too.²⁹

Even though a typical applied researcher does not have access to an experimental estimate and hence cannot benchmark her RDD estimate to the experimental one, it is of interest to ask whether the experimental estimate (Table 2, specification (1)) is statistically different from the non-experimental estimates that the local linear RDD with optimal bandwidths produce (Table 3, specifications (1) and (2)). The reason is that an alternative interpretation for our findings is that our experimental estimate is imprecise and, in fact, consistent with a small and positive incumbency effect. The experimental estimate (0.004) is 88.6% smaller than the RDD estimate (0.039) produced by the local linear RDD with the IK bandwidth, but we cannot reject the null hypothesis that the two estimates are equal (p -value with clustering = 0.23). However, the difference is statistically significant at the 10% level when the RDD estimate based on the CCT bandwidth

²⁷The IK and CCT bandwidths are quite close to each other and they give similar results. For example, the IK bandwidth corresponds to 0.54% of the total votes of a list (that is 5.4 votes out of 1000). This typically translates into a small number of votes. However, the bandwidths are not that small when compared to the vote shares that the candidates at the cutoff get (6.5% vote share, see Table 1). We use here only the CCT bandwidth selection criteria, but not yet the bias-correction or robust inference method that Calonico, Cattaneo, and Titiunik (2014a) also propose, that is, CCT correction.

²⁸Obviously, in some other applications, especially if there are less data available, the bias–variance trade-off could result in larger bandwidths being the preferred approach.

²⁹Card et al. (2014) propose selecting the order of the local polynomial by minimizing the asymptotic MSE. We have used polynomials of orders 0–5 with the IK bandwidths optimized separately for each polynomial specification. We failed to reproduce the experimental estimate using this procedure.

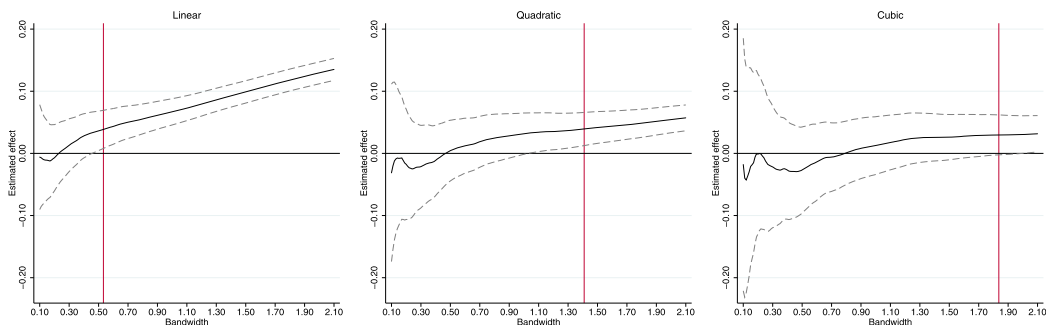


FIGURE 2. Conventional RDD estimates using various bandwidths. The figure displays point estimates from local polynomial regressions with triangular kernel using various bandwidths. Dashed lines show 95% confidence intervals computed using standard errors clustered at the municipality level. Vertical line marks the IK bandwidth.

is used (p -value with clustering = 0.08).³⁰ It is important to stress that this comparison is *not* what a typical applied researcher using RDD absent the experiment could do.

The graphical evidence in Figure 1 suggested that the difference in the estimates is due to the conventional RDD implementation not being able to capture the curvature of the regression function rather than just due to statistical uncertainty. To analyze this further, Figure 2 displays RDD estimates for a large number of bandwidths using the three local polynomial regressions. The vertical bars indicate the location of the optimal bandwidth, which varies with the order of the polynomial. The figure provides us with two main findings. First, the bias relative to the experimental benchmark estimate of 0.004 seems to be almost monotonic in the size of the bandwidth. The approximation gets worse, as more and more data are included in the RDD sample. Even in the absence of the experimental estimate, this finding suggests that there is a need to go beyond a local linear polynomial (or to use a bias correction; see below). This further illustrates the importance of taking the curvature of the regression function into account. Second, when bandwidths narrower than the optimal ones are used, RDD no longer rejects the null hypothesis of no personal incumbency advantage. The null hypothesis is not rejected for the narrower bandwidths both because the point estimate gets smaller and because the confidence intervals get wider.

3.2.4 Bias-corrected RDD estimations Calonico, Cattaneo, and Titiunik (2014a) recently proposed a procedure for bias-correction and robust inference when implementing RDD. The procedure separates point estimation from inference and its goal is to provide valid inference.³¹ The procedure corrects for a bias in the distributional approximation by recentering and rescaling the conventional t -statistic when calculating the robust confidence intervals. In what follows, we call this procedure *CCT correction*. To evaluate how the procedure works, we report a number of RDD estimates using the CCT

³⁰Inference is similar without clustering.

³¹The procedure does not improve point estimation: The conventional RDD point estimator is consistent and MSE-optimal. The bias-corrected point estimator is consistent, but not MSE-optimal.

correction in Table 4. In this method, a p th order local polynomial is used to estimate the main regression discontinuity (RD) effect, whereas a $(p + 1)$ th order local polynomial is used to estimate the (potential) bias. Table 4 consists of three panels. We report in each panel the bias-corrected estimates so as to see how they change relative to the conventional point estimates (reported earlier in Table 3) as well as the nonclustered and clustered 95% confidence intervals.

In Panel A, we use bandwidths optimized for the linear specification, but report the estimates from linear, quadratic, and cubic local polynomial specifications. For this panel, we choose the bias bandwidth (used to estimate the bias) either by the data-driven method suggested by Calonico, Cattaneo, and Titiunik (2014a; using the default option in the pre-2016 *rdr* Stata package; see Calonico, Cattaneo, and Titiunik (2014b)) or by using the IK implementation. When the bias bandwidth is chosen by the data-driven method of Calonico, Cattaneo, and Titiunik (2014a), the RD effect bandwidth is determined to be MSE-optimal, based on the CCT implementation. When the bias bandwidth is chosen by the IK implementation, so is the RD effect bandwidth. The results of this panel show that the CCT-correction is able to meet the replication standard, in the sense that when the CCT-corrected estimates and standard errors are used, we do not, in general, reject the null hypothesis of no effect. The important exception to this result is the data-driven bias bandwidth calculation suggested by Calonico, Cattaneo, and Titiunik (2014a). It apparently leads to bias bandwidths that are too wide. When the bias and RD effect bandwidths are chosen by the IK implementation, the bandwidths are narrower. In this case, the CCT correction meets the replication standard irrespective of which local polynomial specification is used.

In Panel B, we again report the estimates from linear, quadratic, and cubic local polynomial specifications, but choose the bandwidths differently. We optimize the RD effect bandwidths for the linear specification using the CCT and IK implementations. We then impose the bias bandwidth to be the same as the RD effect bandwidth. This is in line with the recent recommendation of Calonico, Cattaneo, and Farrell (2016a), who argue that this is a natural choice with good (numerical) properties. From the perspective of the point estimate, CCT correction with the same bias and RD effect bandwidth amounts to using the conventional local polynomial approach, but with the twist that the main effect is estimated using a one order higher polynomial specification ($p + 1$) than the specification for which the bandwidth is selected (p); see also Calonico, Cattaneo, and Titiunik (2014a). It follows that the point estimate (but not the confidence interval) is the same in columns (4) and (5) of Table 3 as it is in columns (7) and (8) of Panel B of Table 4. The results in this panel show that when implemented in this way, the CCT correction is able to meet the replication standard.

In Panel C, we use the bandwidths optimized for the quadratic and cubic local specifications. They are chosen as in Panel A. We again find that the CCT correction is able to meet the replication standard, provided that the bias and RD effect bandwidths are chosen by the IK implementation. The data-driven method suggested by Calonico, Cattaneo, and Titiunik (2014a) again seems to lead to a bias bandwidth that is too large.

To explore how the bias-corrected and robust estimates vary with different bandwidths and how the two bandwidth choices interact, we display in Figure 3 the bias-corrected RDD estimates and their robust 95% confidence intervals for a fixed bias

TABLE 4. CCT bias-corrected local polynomial RDD estimates with robust inference: Outcome is elected next election.

Panel A: CCT Procedure, Bandwidths Optimized for Local Linear Specification						
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.030	0.046	0.006	0.021	-0.022	-0.004
95% confidence interval (nonclustered)	[-0.002, 0.061]	[0.021, 0.070]	[-0.040, 0.053]	[-0.016, 0.058]	[-0.089, 0.044]	[-0.056, 0.048]
95% confidence interval (clustered)	[-0.003, 0.063]	[0.020, 0.071]	[-0.041, 0.054]	[-0.017, 0.060]	[-0.088, 0.044]	[-0.056, 0.048]
<i>N</i>	19,648	27,218	19,648	27,218	19,648	27,218
Bandwidth	0.54	0.74	0.54	0.74	0.54	0.74
Bias bandwidth	1.14	3.03	1.14	3.03	1.14	3.03
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT

Panel B: CCT Procedure, Equal Bandwidths Optimized for Local Linear Specification						
	(7)	(8)	(9)	(10)	(11)	(12)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.008	0.022	-0.021	-0.004	-0.034	-0.021
95% confidence interval (nonclustered)	[-0.037, 0.053]	[-0.015, 0.059]	[-0.086, 0.045]	[-0.056, 0.048]	[-0.124, 0.056]	[-0.092, 0.050]
95% confidence interval (clustered)	[-0.037, 0.054]	[-0.016, 0.061]	[-0.086, 0.044]	[-0.056, 0.048]	[-0.118, 0.050]	[-0.090, 0.048]
<i>N</i>	19,648	27,218	19,648	27,218	19,648	27,218
Bandwidth	0.54	0.74	0.54	0.74	0.54	0.74
Bias bandwidth	0.54	0.74	0.54	0.74	0.54	0.74
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT

(Continues)

TABLE 4. *Continued.*

Panel C: CCT Procedure, Bandwidths Optimized for Each Polynomial Specification						
	(13)	(14)	(15)	(16)	(17)	(18)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.030	0.046	0.026	0.042	0.025	0.052
95% confidence interval (nonclustered)	[−0.002, 0.061]	[0.021, 0.070]	[−0.007, 0.059]	[0.019, 0.066]	[−0.012, 0.062]	[0.031, 0.072]
95% confidence interval (clustered)	[−0.003, 0.063]	[0.020, 0.071]	[−0.009, 0.061]	[0.019, 0.066]	[−0.014, 0.064]	[0.031, 0.072]
<i>N</i>	19,648	27,218	54,436	78,469	70,576	112,398
Bandwidth	0.54	0.74	1.41	2.09	1.84	3.98
Bias bandwidth	1.14	3.03	1.49	5.38	1.92	7.90
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT

Note: Table shows estimated incumbency advantage using local polynomial regressions within various bandwidths. CCT correction refers to bias-corrected local polynomial RDD estimates with robust inference. All estimations use a triangular kernel. Confidence intervals without clustering are computed using heteroscedasticity-robust standard errors, and clustered confidence intervals account for clustering at the municipality level. The unit of observation is a candidate i at year t . The IK and CCT bandwidths are two different implementations of the estimation of the MSE-optimal theoretical bandwidth choice.

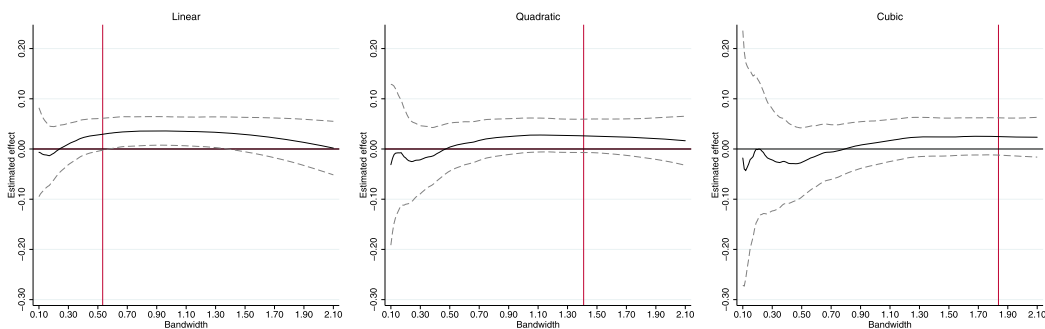


FIGURE 3. Bias-corrected RDD estimates, fixed bias bandwidth. *Notes:* Figure displays bias-corrected point estimates from local polynomial regressions with triangular kernel using various bandwidths. Dashed lines show 95% confidence intervals computed using robust standard errors. Vertical lines mark the IK bandwidth. The bias bandwidth for bias correction has been fixed to 1.14, 1.49 and 1.92 for linear, quadratic and cubic specifications, respectively.

bandwidth, but for different RD effect bandwidths. For this figure, we use the IK implementation to determine the bias bandwidth, because it seemed to lead to narrower bandwidths and worked well. The figure shows that when fixing the bias bandwidth to be IK optimal, the estimated effect is quite robust to the choice of the RD effect bandwidth and most of the time not significantly different from 0.

In Figure 4, we allow both bandwidths to vary and report the corresponding CCT-corrected estimates and their robust confidence intervals. While the results for the linear local polynomial resemble a bit those we reported earlier (Figure 2) for the conventional RDD, there nevertheless is a difference: The figure shows that when the CCT correction is used and the RD effect bandwidth is chosen to be IK optimal or smaller, the null hypothesis of no effect is not rejected in any of the specifications. In line with this, Calonico, Cattaneo, and Farrell (2016a) argue that a bandwidth adjustment (“shrinkage”) is called for to achieve better coverage error rates when MSE-optimal bandwidths are used.

Furthermore, following the recommendation of Calonico, Cattaneo, and Farrell (2016a), we set both bandwidths equal and report the corresponding CCT-corrected estimates and their robust confidence intervals in Figure 5. When the CCT correction is used and the RD effect bandwidth is chosen to be IK optimal or smaller and equal to the bias bandwidth, the null hypothesis of no effect is not rejected. This shows that CCT correction is less sensitive to the choice of the bandwidth (than ad hoc undersmoothing) and works especially well when the bias and RD effect bandwidths are set equal.

In sum, the above findings support the results of Monte Carlo simulations and formal analyses reported in Calonico, Cattaneo, and Titiunik (2014a) and Calonico, Cattaneo, and Farrell (2016a). The above analyses, and especially Figure 5, also support the idea that the CCT bandwidths should be adjusted by a shrinkage factor, as proposed by Calonico, Cattaneo, and Farrell (2016a). Unlike ad hoc undersmoothing, the adjustment improves the coverage error rates of the MSE-optimal bandwidths. For our sample size, the proposed adjustment factors are 0.55, 0.51, and 0.51 for the linear, quadratic, and cubic specifications, respectively. As can be seen from Figure 5, applying the shrinkage

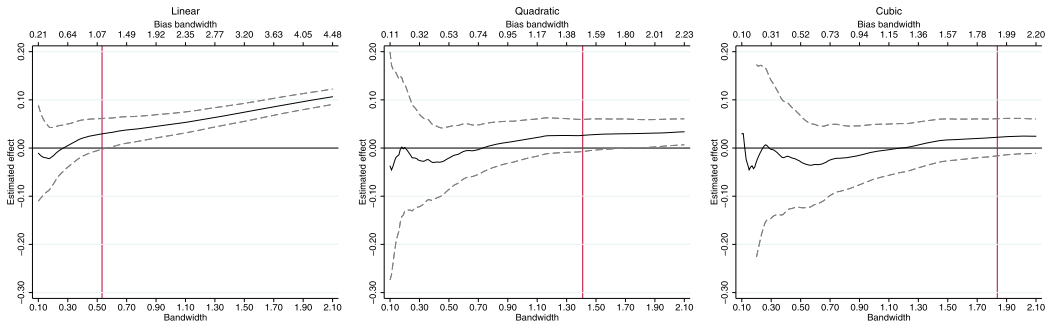


FIGURE 4. Bias-corrected RDD results: both bandwidths vary. The figure displays bias-corrected point estimates from local polynomial regressions with triangular kernel using various bandwidths. Dashed lines show 95% confidence intervals computed using robust standard errors. In the third graph, confidence intervals are omitted for bandwidths smaller than 0.2. Vertical lines mark the IK RD effect and bias bandwidths (both for estimation and bias correction).

factors moves the CCT bandwidth closer to the IK bandwidth and reproduces the experimental result of not rejecting the null hypothesis of no effect. Had the adjustment not been done, the result would have been different.³²

We believe that the above results are useful and of interest to applied econometricians, because most of the existing published and ongoing work applying RDD uses the same implementations of the MSE-optimal bandwidth and bias-correction approach as

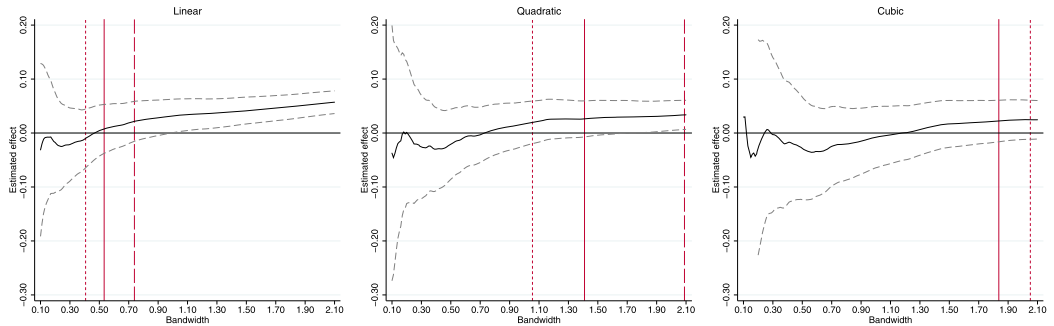


FIGURE 5. Bias-corrected RDD results: bandwidths equal. The figure displays bias-corrected point estimates from local polynomial regressions with triangular kernel using various bandwidths. Dashed lines show 95% confidence intervals computed using robust standard errors. In the third graph, confidence intervals are omitted for bandwidths smaller than 0.2. Vertical solid lines mark the IK bandwidth, long-dashed lines mark the CCT bandwidth, and short-dashed lines mark the adjusted CCT bandwidth. For the cubic specification, the nonadjusted CCT bandwidth does not fit within the x -axis.

³²To keep the graphs comparable, we have not drawn the vertical line for the unadjusted CCT bandwidth of the cubic specification (on the right). The bandwidth is 3.98, leading to a point estimate of 0.035 (with 95% confidence intervals of [0.009, 0.060] with and without clustering at the municipality level).

we have so far done. We analyze some more recent developments briefly in a robustness test reported in the next section.

4. DISCUSSION AND ROBUSTNESS

4.1 *RDD falsification and smoothness tests*

The standard pattern of validity tests for the RDD includes the McCrary (2008) manipulation test—covariate balance tests, which are an indirect test of the smoothness assumption—and placebo tests, where the location of the cutoff is artificially redefined. We do not report the results of the validity tests in detail here. It suffices to note the following (see Appendix D for details).

First, there is no jump in the amount of observations at the cutoff of getting elected. Second, when testing for covariate balance, we allow for the possibility that the covariates have slopes (or even curvature) near the cutoff (e.g., Snyder, Folke, and Hirano (2015) and Eggers et al. (2015)) and estimate local polynomial specifications. We calculate the optimal bandwidths (and half the optimal ones) for different polynomials to address potential slope and curvature issues. We do this for each covariate separately. The covariate balance tests produce somewhat mixed evidence, but overall they suggest that RDD ought to work well in our application. This finding is somewhat in contrast with those of Caughey and Sekhon (2011), who mention the possibility that purposeful sorting by the candidates may also invalidate the use of RDD in the closest races. We find some evidence that there are fewer rejections of covariate balance when more flexible local polynomial specifications (or undersmoothing) are used.

Finally, the placebo cutoff tests provide signals that cast doubt on the appropriateness of standard local linear (and polynomial) RDD specifications with the MSE-optimal bandwidths in our context. Moreover, the placebo tests do not suggest that undersmoothing procedures and use of higher degree local polynomials without adjusting the bandwidth accordingly would not work. This finding echoes the conclusion that when these bias-correction tools are used, RDD is able to reproduce the experimental estimate. In sum, this shows that the placebo cutoff tests can be useful in detecting specifications that are too inflexible.

4.2 *Robustness of RDD estimates*

We have conducted a large number of auxiliary analyses and tests to probe the robustness of our RDD findings. We take each of them in turn (see Appendix E for details).

First, RDD is sometimes implemented using higher order global polynomials of the forcing variable. We have redone the RDD analysis using such parametric RDDs, using five different polynomials (1st–5th degree). These parametric RDD generate positive and statistically significant incumbency effects that are roughly similar in magnitude to those reported in Lee (2008). Consistent with what Gelman and Imbens (2014) argue, we find that this approach to implementing RDD provides misleading findings. The bias here is an order of magnitude larger than in the local polynomial specifications.

Second, we have considered the vote share in the subsequent elections as an alternative measure of incumbency advantage. As we reported earlier, the experimental estimate also suggests no incumbency advantage when this alternative measure is used. The pattern of RDD estimates is similar that observed for the main outcome, but none of the reported specifications are significant.

Third, ties appear a bit more often in elections in the smaller municipalities. As we reported earlier, the experimental estimate is quite precisely estimated and close to zero both in small and in large elections. However, our normalized forcing variable can get values really close to zero only when parties get a large amount of votes, which tends to happen in the elections in the larger municipalities. To check what this implies for our RDD findings, we have rerun parts of the RDD analysis separately for small and large municipalities. These estimations show that for both the larger and smaller municipalities, the bias increases with the bandwidth and decreases as the degree of local polynomial increases. It thus seems that our conclusions are not driven by the size of the municipalities.

Fourth, another potential explanation for why the local linear RDD point estimates increase is heterogeneity in the personal incumbency effect across municipalities (and party lists).³³ To explore how much this kind of heterogeneity matters, we have repeated the RDD analysis using only those party lists that were involved in the lotteries. In this case, increasing the bandwidth adds new candidates from the same lists, but does not add new lists or municipalities. Our main results remain unchanged. This analysis is also important because it guarantees that the same set of within-party cutoffs is used in both the experimental sample and the RD sample.

Fifth, we have rerun the RDD estimations using alternative definitions for the forcing variable. The results show that our RDD findings are not driven by the choice of the forcing variable. For example, we get very similar results if the forcing variable is either the vote margin that is calculated in terms of the number of votes or vote shares.

Sixth, we have studied whether there is heterogeneity in the effect between the parties. We found no evidence for substantial heterogeneity in the personal incumbency advantage between the parties.

Seventh, we have already mentioned that the experimental estimate does not change if those who do not rerun are excluded from the lottery sample. We have replicated our baseline RDD analysis using the sample from which those who do not rerun are similarly excluded. Our results remain unchanged.

Finally, we want to acknowledge that a major upgrade of `rdrobust` software is now available (Calonico et al. (2016b)). The updated version introduces a new implementation of the MSE-optimal bandwidth choice, replacing the IK and CCT implementations. The new implementation of the MSE-optimal bandwidth estimates the same asymptotic quantity as the CCT and IK implementations. The updated software allows for clustering when calculating standard errors and bandwidths. We have reestimated the most relevant specifications of the previous sections using the new implementation of the

³³ Changing the bandwidth used for estimation does not change the parameter that is being identified. When the width of the bandwidth is changed, the accuracy of the approximation used to estimate the parameter changes.

MSE-optimal bandwidth with and without clustering at municipality level. The results largely echo our earlier findings (Appendix E). In particular, because the new implementation of the MSE-optimal bandwidth is similar to the CCT implementation, the results look alike.³⁴

4.3 *When is RDD as good as random?*

One reason for the popularity of RDD is that close to the cutoff, variation in the treatment status may be as good as random, provided that the forcing variable cannot be precisely manipulated (Lee (2008, p. 676)). RDD is widely believed to meet the replication standard because of this feature. While somewhat distinct from our previous analysis, this naturally leads to the question of whether we can identify a neighborhood around the cutoff where the randomization assumption is plausible (Cattaneo, Frandsen, and Titiunik (2015)). To answer the question, we explore the largest bandwidth in which the as-good-as-random assumption holds and then compare the sample means of the outcome variable across the cutoff.³⁵ We find that (see Appendix F), with some caveats, we can reproduce the experimental estimate using the approach proposed by Cattaneo, Frandsen, and Titiunik (2015).

5. CONCLUSIONS

We have made use of elections data in which the electoral outcome was determined via a random seat assignment for a large number of candidates because of a tie in their vote count. These instances provide us with a randomized experiment against which we have benchmarked non-experimental RDD estimates of personal incumbency advantage. To our knowledge, the experiment is unique in the literature because it takes place exactly at the cutoff. This means that both the experiment and the RDD target the same treatment effect.³⁶

³⁴Moreover, the update introduces the so-called coverage-error-rate -optimal (CER-optimal) bandwidth, which is a bandwidth choice based on a higher order Edgeworth expansion (Calonico et al. (2016b)). This bandwidth optimizes coverage error but does not necessarily have desirable properties for point estimation. The results based on the CER-optimal bandwidth also echo our earlier findings (Appendix E).

³⁵As Cattaneo, Frandsen, and Titiunik (2015), Cattaneo, Titiunik, and Vazquez-Bare (2017), de la Cuesta and Imai (2016), and Skovron and Titiunik (2015) have emphasized, the (local) randomization assumption differs from the usual assumption of no discontinuity in the conditional expectation function of the potential outcome. This randomization feature of RDD may be the reason why RDD has been used as a benchmark against which other non-experimental estimators have been compared (see, e.g., Lemieux and Milligan (2008)). We know that in a sample that only includes the lotteries (i.e., when the neighborhood is degenerate at the cutoff), the randomization assumption is satisfied in our data. The subsample that we use to explore the plausibility of the randomization assumption excludes the randomized candidates.

³⁶To be precise, this statement is true if the cutoff were the same for all observations. In our application, there are multiple cutoffs that are all normalized to 0. As Cattaneo, Titiunik, and Vazquez-Bare (2016, 2017) explain, the pooled RDD estimand over multiple cutoffs depends on two things. First, it depends on the density of observations at the individual cutoffs. Second, the estimand is a function of the probability of each observation facing a given cutoff value. In our robustness tests, we restrict the sample so that the cutoff is the same for all observations in the estimation. Our main findings are robust in this regard.

We find no evidence of a personal incumbency advantage when the data from the randomized elections are used. The point estimate of the incumbency advantage is close to 0 and relatively precisely estimated. We argue that this finding is neither surprising nor in conflict with the prior evidence, because we are looking at the effect of incumbency status on electoral success in a rather special context, in small local PR elections. It is possible that the randomized electoral victories as well as the close elections that we study take place at a relatively unimportant margin, providing limited scope for the emergence and creation of personal incumbency advantage.

Our two main RDD findings are the following. First, when RDD is applied in conventional fashion (i.e., using local linear regression with MSE-optimal bandwidths) to the same close elections, the estimates suggest a moderate and statistically significant personal incumbency effect. Second, the recent bias correction and robust inference method of Calonico, Cattaneo, and Titiunik (2014a) is able to recover the experimental benchmark, provided that bias bandwidths that are not too wide are used. We find that the procedure is less sensitive to the choice of the bandwidth (than ad hoc undersmoothing) and works especially well when the bias and RD effect bandwidths are set equal. These results are important, because compared to the often-used alternative approach of undersmoothing, the method of Calonico, Cattaneo, and Titiunik (2014a) is more efficient and has faster vanishing coverage error rates. Our findings corroborate the findings of the simulation and formal analyses of Calonico, Cattaneo, and Titiunik (2014a) and Calonico, Cattaneo, and Farrell (2016a), which demonstrate that the method of Calonico, Cattaneo, and Titiunik (2014a) ought to work better than the conventional ad hoc adjustments.

These findings lead to two key conclusions. First, RDD can indeed meet the replication standard in the context of close elections. Second, and more interestingly, the results may be sensitive to the details of implementation even when the researcher has a relatively large number of observations. The recently proposed implementation approaches work better than the older ones.

REFERENCES

- Bartalotti, O. C. and Q. O. Brummet (2017), “Regression discontinuity designs with clustered data: Variance and bandwidth choice.” In *Regression Discontinuity Designs: Theory and Applications* (M. D. Cattaneo and J. C. Escanciano, eds.), *Advances in Econometrics*, Vol. 38, 383–420, Emerald Publishing Limited. [1033]
- Black, D. A., J. Galdo, and J. A. Smith (2007), “Evaluating the worker profiling and reemployment services system using a regression discontinuity approach.” *American Economic Review, Papers and Proceedings*, 97 (2), 104–107. [1020, 1022]
- Broockman, D. E. (2009), “Do congressional candidates have reverse coattails? Evidence from a regression discontinuity design.” *Political Analysis*, 17 (4), 418–434. [1029]
- Calonico, S., M. D. Cattaneo, M. F. Farrell, and R. Titiunik (2016), “*rdrobust*: Software for regressions discontinuity designs.” *Stata Journal*, 17 (2), 372–404. [1044, 1045]

- Calonico, S., M. D. Cattaneo, and M. F. Farrell (forthcoming), “On the effect of bias estimation on coverage accuracy in non-parametric inference.” *Journal of the American Statistical Association*. [1021, 1022, 1038, 1041, 1046]
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a), “Robust nonparametric confidence intervals for regression-discontinuity designs.” *Econometrica*, 82 (6), 2295–2326. [1021, 1022, 1031, 1036, 1037, 1038, 1041, 1046]
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014b), “Robust data-driven inference in the regression discontinuity design.” *Stata Journal*, 14 (4), 909–946. [1031, 1038]
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2015), “Optimal data-driven regression discontinuity plots.” *Journal of the American Statistical Association*, 110 (512), 1753–1769. [1031, 1032]
- Card, D., D. S. Lee, Z. Pei, and A. Weber (2014), “Local polynomial order in regressions discontinuity designs.” Brandeis University Working Paper 81. [1036]
- Cattaneo, M. D., B. R. Frandsen, and R. Titiunik (2015), “Randomization inference in the regression discontinuity design: An application to party advantages in the U.S. senate.” *Journal of Causal Inference*, 3 (1), 1–24. [1023, 1045]
- Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare (2016), “Inference in regression discontinuity designs under local randomization.” *Stata Journal*, 16 (2), 331–367. [1045]
- Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare (2017), “Comparing inference approaches for rd designs: A reexamination of the effect of head start on child mortality.” *Journal of Policy Analysis and Management*, 36 (3), 643–681. [1045]
- Caughey, D. and J. S. Sekhon (2011), “Elections and the regression discontinuity design: Lessons from close U.S. House races, 1942–2008.” *Political Analysis*, 19 (4), 385–408. [1023, 1029, 1043]
- Cook, T. D., W. Shadish, and V. C. Wong (2008), “Three conditions under which observational studies produce the same results as experiments.” *Journal of Policy Analysis and Management*, 27 (4), 724–750. [1022, 1023]
- Cook, T. D. and V. C. Wong (2008), “Empirical tests of the validity of the regression discontinuity design.” *Annals of Economics and Statistics*, 91–92, 127–150. [1022]
- Cox, G. W. and J. N. Katz (1996), “Why did the incumbency advantage in U.S. House elections grow?” *American Journal of Political Science*, 40 (2), 478–497. [1029]
- Dahlgard, J. O. (2016), “You just made it: Individual incumbency advantage under proportional representation.” *Electoral Studies*, 44, 319–328. [1023, 1030]
- de la Cuesta, B. and K. Imai (2016), “Misunderstandings about the regression discontinuity design in the study of close elections.” *Annual Review of Political Science*, 19, 375–396. [1020, 1045]
- De Magalhaes, L. (2014), “Incumbency effects in a comparative perspective: Evidence from Brazilian mayoral elections.” *Political Analysis*, 23 (1), 113–126. [1020]

- Dehejia, R. and S. Wahba (2002), “Propensity score-matching methods for nonexperimental causal studies.” *Review of Economics and Statistics*, 84 (1), 151–161. [1020]
- Desposato, S. W. and J. R. Petrocik (2003), “The variable incumbency advantage: New voters, redistricting, and the personal vote.” *American Journal of Political Science*, 47 (1), 18–32. [1029]
- Eggers, A. C., A. Fowler, J. Hainmueller, A. B. Hall, and J. M. Snyder (2015), “On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races.” *American Journal of Political Science*, 59 (1), 259–274. [1023, 1043]
- Erikson, R. S. and K. Rader (2017), “Much ado about nothing: rdd and the incumbency advantage.” *Political Analysis*, 25 (2), 269–275. [1023]
- Erikson, R. S. and R. Titiunik (2015), “Using regression discontinuity to uncover the personal incumbency advantage.” *Quarterly Journal of Political Science*, 10 (1), 101–119. [1020, 1029, 1030]
- Fan, J. and I. Gijbels (1996), *Local Polynomial Modeling and Its Applications*. Chapman & Hall/CRC Press, New York. [1031]
- Ferreira, F. and J. Gyourko (2009), “Do political parties matter? Evidence from U.S. cities.” *Quarterly Journal of Economics*, 124 (1), 399–422. [1020]
- Folke, O. (2014), “Shades of brown and green: Party effects in proportional election systems.” *Journal of the European Economic Association*, 12 (5), 1361–1395. [1025, 1030]
- Folke, O. and J. M. Snyder (2012), “Gubernatorial midterm slumps.” *American Journal of Political Science*, 56 (4), 931–948. [1020]
- Fowler, A. and A. B. Hall (2014), “Disentangling the personal and partisan incumbency advantages: Evidence from close elections and term limits.” *Quarterly Journal of Political Science*, 9 (4), 501–531. [1029]
- Fraker, T. and R. Maynard (1987), “The adequacy of comparison group designs for evaluations of employment-related programs.” *Journal of Human Resources*, 22 (2), 194–227. [1020]
- Gelman, A. and G. Imbens (2014), “Why higher order polynomials should not be used in regression discontinuity designs.” Working Paper 20405, National Bureau of Economic Research. [1043]
- Gelman, A. and G. King (1990), “Estimating incumbency advantage without bias.” *American Journal of Political Science*, 34 (4), 1142–1164. [1029]
- Gerber, E. R. and D. J. Hopkins (2011), “When mayors matter: Estimating the impact of mayoral partisanship on city policy.” *American Journal of Political Science*, 55 (2), 326–339. [1020]
- Golden, M. and L. Picci (2015), “Incumbency effects under proportional representation: Leaders and backbenchers in the postwar Italian chamber of deputies.” *Legislative Studies Quarterly*, 40 (4), 509–538. [1023, 1030]

- 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. [1022]
- Grimmer, J., E. Hersh, B. Feinstein, and D. Carpenter (2011), "Are close elections random?" Working paper. [1023, 1029]
- Hahn, J., P. Todd, and W. van der Klaauw (2001), "Identification and estimation of treatment effects with a regression-discontinuity design." *Econometrica*, 69 (1), 201–209. [1021]
- Iltasanomat (2011), "Tiesitkö tätä vaaleista? Näin käy, jos äänet menevät tasan." 4/12/2011. Available at <http://www.iltasanomat.fi/vaalit2011/art-1288381910312.html>. [1024]
- Imbens, G. and K. Kalyanaraman (2012), "Optimal bandwidth choice for the regression discontinuity estimator." *Review of Economic Studies*, 79 (3), 933–959. [1021, 1031]
- Imbens, G. and T. Lemieux (2008), "Regression discontinuity designs: A guide to practice." *Journal of Econometrics*, 142 (2), 615–635. [1020, 1021, 1031]
- Klašnja, M. and R. Titunik (2016), "The incumbency curse: Weak parties, term limits, and unfulfilled accountability." *American Political Science Review*, 111 (1), 129–148. [1026, 1029]
- Kotakorpi, K., P. Poutvaara, and M. Terviö (2017), "Returns to office in national and local politics: A bootstrap method and evidence from Finland." *Journal of Law, Economics, and Organization*, 33 (3), 413–442. [1023, 1030]
- LaLonde, R. J. (1986), "Evaluating the econometric evaluations of training programs with experimental data." *American Economic Review*, 76 (4), 604–620. [1020, 1022]
- Lee, D. S. (2008), "Randomized experiments from non-random selection in U.S. House elections." *Journal of Econometrics*, 142 (2), 675–697. [1020, 1021, 1029, 1031, 1043, 1045]
- Lee, D. S. and T. Lemieux (2010), "Regression discontinuity designs in economics." *Journal of Economic Literature*, 48 (2), 281–355. [1031]
- Lee, D. S., E. Moretti, and M. J. Butler (2004), "Do voters affect or elect policies? Evidence from the U.S. House." *Quarterly Journal of Economics*, 119 (3), 807–859. [1020]
- Lemieux, T. and K. Milligan (2008), "Incentive effects of social assistance: A regression discontinuity approach." *Journal of Econometrics*, 142, 807–828. [1045]
- Ludwig, J. and D. L. Miller (2007), "Does head start improve children's life chances? Evidence from a regression discontinuity design." *Quarterly Journal of Economics*, 122 (1), 159–208. [1021, 1031]
- Lundqvist, H. (2013), "Is it worth it? On the returns to holding political office." IEB Working paper 2013/14. [1023, 1030]
- McCrary, J. (2008), "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics*, 142 (2), 698–714. [1043]

- Redmond, P. and J. Regan (2015), “Incumbency advantage in a proportional electoral system: A regression discontinuity analysis of Irish elections.” *European Journal of Political Economy*, 38, 244–256. [1023]
- Shadish, W. R., R. Galindo, V. C. Wong, P. M. Steiner, and T. D. Cook (2011), “A randomized experiment comparing random to cutoff-based assignment.” *Psychological Methods*, 16 (2), 179–191. [1022, 1023]
- Skovron, C. and R. Titiunik (2015), “A practical guide to regression discontinuity designs in political science.” Working paper. [1020, 1045]
- Smith, J. and P. E. Todd (2005), “Does matching overcome LaLonde’s critique of nonexperimental estimators?” *Journal of Econometrics*, 125, 305–353. [1020]
- Snyder, J. (2005), “Detecting manipulation in U.S. House elections.” Working paper. [1023]
- Snyder, J. M., O. Folke, and S. Hirano (2015), “Partisan imbalance in regression discontinuity studies based on electoral thresholds.” *Political Science Research and Methods*, 3 (2), 169–186. [1025, 1043]
- Stone, P. (2011), *The Luck of Draw: The Role of Lotteries in Decision-Making*. Oxford University Press. [1020]
- The Atlantic (2012), “When a state election can be literally determined by a coin toss.” 11/19/2012. Available at <http://www.theatlantic.com/politics/archive/2012/11/when-a-state-election-can-be-literally-determined-by-a-coin-toss/265413/>. [1020]
- The Telegraph India (2014), “In election tie, pick winner by lottery.” 2/7/2014. Available at http://www.telegraphindia.com/1140207/jsp/nation/story_17909869.jsp#.VCUt8hCfjv. [1021]
- Thistlethwaite, D. L. and D. T. Campbell (1960), “Regression-discontinuity analysis: An alternative to the ex post facto experiment.” *Journal of Educational Psychology*, 51 (6), 309–317. [1020]
- Time (2013), “Coin toss determines mayor in Philippine town.” 5/17/2013. Available at <http://newsfeed.time.com/2013/05/17/coin-toss-determines-mayor-in-philippine-town/>. [1020]
- Trounstine, J. (2011), “Evidence of a local incumbency advantage.” *Legislative Studies Quarterly*, 36 (2), 255–280. [1029]
- UPI (2014), “New Mexico judicial election decided by coin toss.” 7/14/2014. Available at http://www.upi.com/Top_News/US/2014/07/14/New-Mexico-judicial-election-decided-by-coin-toss/6161405361068/. [1020]
- Uppal, Y. (2009), “The disadvantaged incumbents: Estimating incumbency effects in Indian state legislatures.” *Public Choice*, 138 (1), 9–27. [1029]
- Wing, C. and T. D. Cook (2013), “Strengthening the regression discontinuity design using additional design elements: A within-study comparison.” *Journal of Policy Analysis and Management*, 32 (4), 853–877. [1022]

Co-editor Peter Arcidiacono handled this manuscript.

Manuscript received 3 April, 2017; final version accepted 26 August, 2017; available online 25 October, 2017.