Ari Hyytinen, Jaakko Meriläinen, Tuukka Saarimaa, Otto Toivanen and Janne Tukiainen

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

Article (Accepted version) (Refereed)

Original citation:
Hyytinen, Ari and Meriläinen, Jaakko and Saarimaa, Tuukka and Toivanen, Otto and Tukiainen, Janne (2017) *When does regression discontinuity design work? Evidence from random election outcomes.* Quantitative Economics. ISSN 1759-7323

When Does Regression Discontinuity Design Work? Evidence from Random Election Outcomes

Ari Hyytinen, Jaakko Meriläinen, Tuukka Saarimaa, Otto Toivanen and Janne Tukiainen*

Abstract: 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 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 applying 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 codes: C21, C52, D72.

---

* Hyytinen: University of Jyväskylä, School of Business and Economics, P.O. Box 35, 40014 University of Jyväskylä, Finland, ari.hyytinen@jyu.fi; Meriläinen: Institute for International Economic Studies, Stockholm University, SE-10691, Stockholm, Sweden, jaakko.merilainen@iies.su.se; Saarimaa: VATT Institute for Economic research, Arkadiankatu 7, Helsinki, FI-00101, tuukka.saarimaa@vatt.fi; Toivanen: Aalto University School of Business and KU Leuven, Arkadiankatu 7, Helsinki, FI-00101, otto.toivanen@aalto.fi; Tukiainen (corresponding author): Department of Government, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK, and VATT Institute for Economic research, Arkadiankatu 7, Helsinki, FI-00101, 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, 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 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

[1] Investigating 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 et al. 2004, Ferreira and Gyourko 2009, Gerber and Hopkins 2011, Folke and Snyder 2012, De Magalhaes 2014). De la Cuesta and Imai (2016) and Skovron and Titiunik (2015) are recent surveys of the close elections RDD analyses.

[2] Black et al. (2007) come close to our analysis, because their experiment targets a population within a small bandwidth around the cutoff. However, as Black et al. (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".

advantage, i.e., 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 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 et al. (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 Kalyanaram 2012, Calonico et al. 2014a; see also Calonico et al. 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 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

---

[3] Use of lotteries to solve electoral ties is not unique to Finland. For example, some US state elections and many US local elections have used lottery-based rules to break ties in elections (see, e.g., UPI 14.7.2014, The Atlantic 19.11.2012, and Stone 2011). Lotteries have been used to determine the winner in case of ties also in the Philippines (Time 17.5.2013), in India (The Telegraph India 7.2.2014), in Norway as well as in Canada and the UK (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.

[4] Another 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

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 et al. 2014a).

We find that when an *ad hoc* under-smoothing procedure of using smaller (than MSE-optimal) bandwidths is used to reduce the bias (see, e.g., Imbens and Lee 2008; Calonico et al. 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 et al. (2014a) brings the RDD estimate(s) in line with the experimental 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* under-smoothing procedures) that have faster vanishing coverage error rates (see also Calonico et al. 2016a). Given that we build on a real-world experiment, we provide an independent verification of the empirical performance of Calonico et al. (2014a) procedure: We find that the procedure is less sensitive to the choice of the bandwidth (than *ad hoc* under-smoothing) 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 et al. (2014a) and Calonico et al. (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

---

not continuous at the cutoff. We find no signs of this key RDD assumption being violated using covariate balance checks.

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

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 et al. (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 et al. (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 et al. (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, 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 et al. 2007, Cook and Wong 2008, Cook et al. 2008, Green et al. 2009, Shadish et al. 2011, 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. 2012). Eggers et al. (2015) convincingly

[6] The 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 et al. 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.

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 et al. (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 2011, Redmond and Regan 2015, Golden and Picci 2015, Dahlgaard 2016 and Kotakorpi et al. 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 an online appendix that supplements this paper.

2 Institutional context and data

2.1 Institutional environment

Finland has a two-tier system of government, consisting 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 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.

[7] The 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 dataset on 40 000 close political races cover.

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

[9] Municipalities employ around 20 percent 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.

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 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 $(S+1)^{\text{th}}$ (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 $(k+1)^{\text{th}}$ 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 nor implemented a randomized controlled trial, we

[10] See e.g. an article in one of the major Finnish tabloids, Iltasanomat, on 12.4.2011.

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 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 multi-party 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 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 socio-economic 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

[11] See Folke (2014) for the complications that multi-party-systems generate and Snyder et al. (2015) on issues with partisan imbalance in RDD studies.

[12] The 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.

[13] Two 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.

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

In elections between 1996 and 2008, 1351 candidates had a tie within their party lists for the last seat(s), i.e. at the margin which 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*, i.e. among the candidates who had a tie.

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]

[14] In addition, there were 202 ties in 2012. We do not include them in the lottery sample, because we don't 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.

[15] The 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.

Table 1. Covariate balance tests for the lottery sample.

	Elected (N = 671)			Not elected (N = 680)					
Variable	N	Mean	Std. Dev.	N	Mean	Std. Dev.	Difference	p-value	p-value (clustered)
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	22521	14928	476	22256	13729	265	0.78	0.82
Capital income	478	2929	18612	476	3234	12085	-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

Notes: Difference in means has been tested using *t*-test with and without clustering at municipality level. 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 expressed in nominal euros.

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 as 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

[16] In the party level analysis of Klasnja and Titiunik (2016), the dependent variable is a binary variable equal to 1 if the party wins the election at $t + 1$, and is equal to zero 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). Klasnja and Titiunik (2016) also report an analysis conditioning on running again in an appendix.

variable so that it is equal to one if the candidate is elected in the next election, and is set to zero 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, this ATE is estimated *precisely at the cutoff point* of political support which 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 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 zero. The estimate is insignificant also if a conventional (non-robust, non-clustered) 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 zero across the columns and the 95% confidence intervals always include zero. 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.

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

Notes: Only actual lotteries are included in the regressions. Set of controls includes age, gender, party affiliation, socio-economic 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 municipality level. Unit of observation is a candidate *i* at year *t*.

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 zero 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 zero (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.

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, 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 the references therein). It is clear that the size of the advantage may nevertheless vary and be context specific; see e.g. Desposato and Petrocik (2003), Grimmer et al. (2012), 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's 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 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.

[17] The 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, 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, Trounstine 2011; see also Fowler and Hall 2014).

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

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 the prior work. Moreover, the existing studies that look at a personal incumbency advantage in the PR systems of developed countries find typically only modest or no incumbency effects (Lundqvist 2011, Golden and Picci 2015, Dahlgaard 2016 and Kotakorpi et al. 2017).

3.2 Non-experimental estimates

Implementing RDD for PR elections

Our forcing variable is constructed as follows. We measure closeness *within a party list* in order to focus on the same cutoff where the lotteries take place, and to abstract from multi-party 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

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

[20] This definition of the forcing variable means that all those party lists from which no candidates or all candidates got elected are dropped out 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.

[21] Dahlgaard (2016), Golden and Picci (2015), Lundqvist (2011) and Kotakorpi et al. (2017) study quasi-randomization that takes place within parties in a PR system using an approach similar to ours

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 (potential alternative) forcing variable based on the absolute number of votes operate on a very different scale, but they are correlated (their pairwise correlation is in our data 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. 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, 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) and Calonico et al. (2014a, CCT).[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.

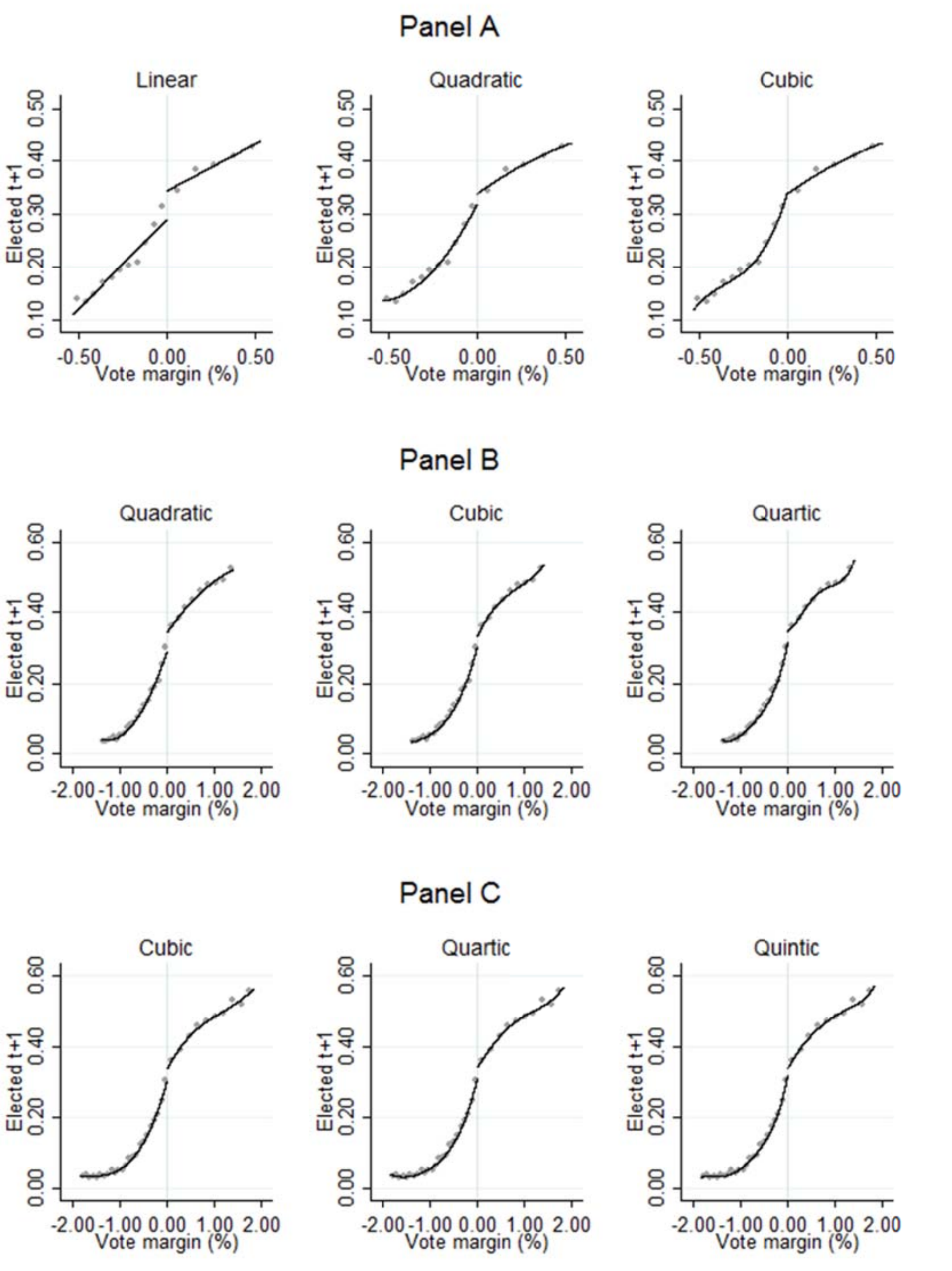
[22] In large elections, it is more likely that small vote share differences are observed (rather than small differences in the number of votes). The opposite holds for small elections.

[23] Two 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 et al. 2014a and 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.

RDD estimations: Graphical analysis

We start by displaying the relation 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 one 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 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 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 two might be flexible enough for the bandwidth that has been optimized for a polynomial of order one.

[24] The figure has been produced by the rdplot command for Stata that approximates the underlying unknown regression functions without imposing smoothness (Calonico et al. 2015). The key contribution in Calonico et al. (2015) is to provide a data driven approach for choosing the bin widths which 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.



Notes: 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, quadratic specification in Panel B and 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 the optimal bandwidths are calculated for. 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 IMSE-optimal evenly-spaced approach of Calonico et al. (2015).

Figure 1. Curvature between the forcing variable and the outcome.

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. Like 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 non-existent in the graphs displayed in the middle (*p*+1) or on the right (*p*+2).[25]

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 CCT implementation of the bandwidth selection. The panel reports for these bandwidth choices the local linear (specifications (1)-(2)), quadratic (specifications (3)-(4)) and cubic (specifications (5)-(6)) RDD estimates of the personal incumbency advantage. As specifications (1)-(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.

[25] We 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).

[26] We 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 (2016) for a recent analysis of cluster-based inference in the context of RDD.

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, 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 under-smoothing 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)-(8)). The null hypothesis of no personal incumbency effect cannot be rejected either 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 over-rejection because the distributional approximation of the estimator is poor (Calonico et al. 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 et al. 2014a for a discussion of under-smoothing).[28] Moreover, we find that holding the bandwidth constant, richer polynomials align our RDD results with the experimental benchmark, too.[29]

---

[27] The IK and CCT bandwidths are quite close to each other and they give similar results. For example, the IK bandwidth corresponds to 0.53% of the total votes of a list (that is 5.3 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 et al. (2014a) also propose, i.e., CCT-correction.

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

[29] Card 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.

Table 3. Local polynomial RDD estimates.

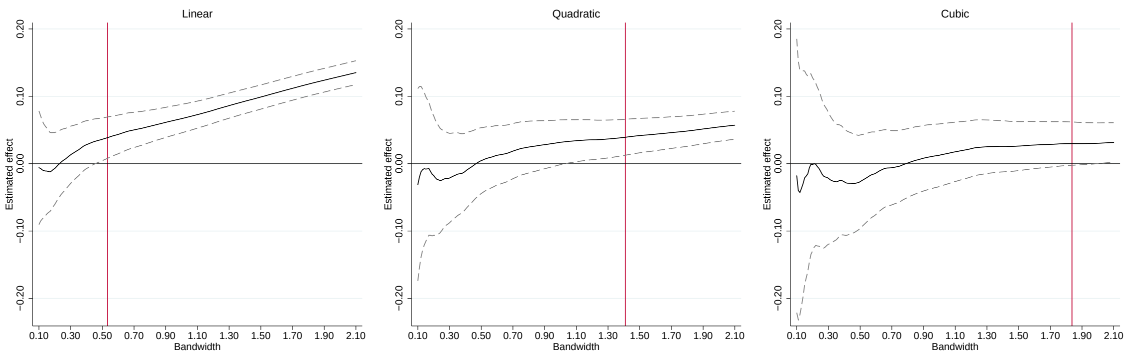
Panel A: Bandwidth optimized for local linear specification						
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear		Quadratic		Cubic	
Elected	0.039	0.052	0.008	0.022	-0.022	-0.004
95% confidence interval (robust)	[0.010, 0.068]	[0.028, 0.076]	[-0.037, 0.053]	[-0.015, 0.059]	[-0.087, 0.043]	[-0.056, 0.048]
95% confidence interval (clustered)	[0.008, 0.069]	[0.027, 0.077]	[-0.039, 0.055]	[-0.039, 0.055]	[-0.088, 0.044]	[-0.058, 0.049]
N	19407	26999	19407	26999	19407	26999
R^2	0.03	0.05	0.03	0.05	0.03	0.05
Bandwidth	0.53	0.74	0.53	0.74	0.53	0.74
Bandwidth implementation	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	0.007	0.024	-0.022	-0.015	-0.018	-0.025
95% confidence interval (robust)	[-0.036, 0.050]	[-0.011, 0.059]	[-0.095, 0.050]	[-0.073, 0.043]	[-0.128, 0.092]	[-0.111, 0.061]
95% confidence interval (clustered)	[-0.039, 0.044]	[-0.013. 0.061]	[-0.094, 0.049]	[-0.074, 0.045]	[-0.126, 0.090]	[-0.108, 0.058]
N	9808	13496	9808	13496	9808	13496
R^2	0.01	0.02	0.01	0.02	0.02	0.02
Bandwidth	0.27	0.37	0.27	0.37	0.27	0.37
Bandwidth implementation	IK	CCT	IK	CCT	IK	CCT
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)	[0.010, 0.068]	[0.028, 0.076]	[0.013, 0.065]	[0.036, 0.078]	[-0.000, 0.060]	[0.035, 0.076]
95% confidence interval (clustered)	[0.008, 0.069]	[0.027, 0.077]	[0.013, 0.066]	[0.036, 0.078]	[-0.002, 0.062]	[0.035, 0.076]
N	19407	26999	54464	78469	70576	112398
R^2	0.03	0.05	0.11	0.16	0.15	0.23
Bandwidth	0.53	0.74	1.41	2.09	1.84	3.98
Bandwidth implementation	IK	CCT	IK	CCT	IK	CCT

Notes: 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. 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.

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)-(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 = 0.24). However, the difference is statistically significant at 10% level when the RDD estimate based on the CCT bandwidth is used (*p*-value with clustering = 0.09).[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.

[30] Inference is similar without clustering.



Notes: 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. Red vertical line marks the IK bandwidth.

Figure 2. Conventional RDD estimates using various bandwidths.

Bias-corrected RDD estimations

Calonico et al. (2014a) have 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 re-centering and re-scaling 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-correction in Table 4. In this method, a p^{th} order local polynomial is used to estimate the main 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 in order to see how they change relative to the conventional point estimates, reported earlier in Table 3, as well as the non-clustered 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 et al. (2014a; using the default option in the pre-2016 rdrobust Stata-package; see Calonico et al. 2014b) or by using the IK implementation. When the bias bandwidth is chosen by the data-driven method of Calonico et al. (2014a), the RD effect

[31] The 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.

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 et al. (2014a). It apparently leads to too wide bias bandwidths. 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, irrespectively 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 et al. (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 et al. (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 of 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 et al. (2014a) again seems to lead to a too large bias bandwidth.

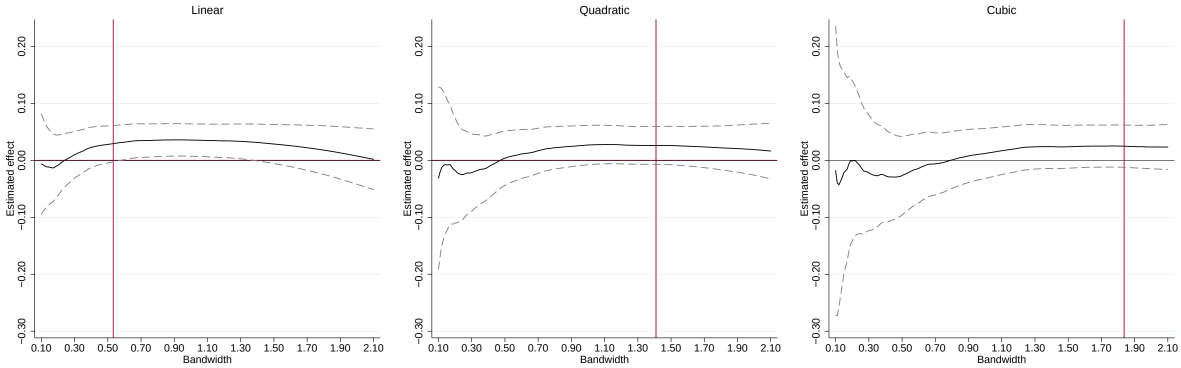
Table 4. CCT bias-corrected local polynomial RDD estimates with robust inference.

Panel A: CCT-correction, bias and RD effect 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	-0.023	-0.004
95% confidence interval (non-clustered)	[-0.001, 0.060]	[0.022, 0.069]	[-0.040, 0.051]	[-0.015, 0.057]	[-0.088, 0.042]	[-0.055, 0.046]
95% confidence interval (clustered)	[-0.004, 0.063]	[0.020, 0.071]	[-0.042, 0.054]	[-0.018, 0.060]	[-0.089, 0.043]	[-0.057, 0.048]
N	19407	26999	19407	26999	19407	26999
RD effect bandwidth	0.53	0.74	0.53	0.74	0.53	0.74
Bias bandwidth	1.14	3.03	1.14	3.03	1.14	3.03
Bandwidth implementation	IK	CCT	IK	CCT	IK	CCT
Panel B: CCT-correction, bias and RD effect bandwidths set equal, optimized for local linear specification						
	(7)	(8)	(9)	(10)	(11)	(12)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.008	0.022	-0.022	-0.004	-0.033	-0.021
95% confidence interval (non-clustered)	[-0.036, 0.051]	[-0.014, 0.057]	[-0.086, 0.042]	[-0.055, 0.046]	[-0.122, 0.056]	[-0.090, 0.048]
95% confidence interval (clustered)	[-0.038, 0.053]	[-0.017, 0.060]	[-0.087, 0.043]	[-0.056, 0.048]	[-0.118, 0.051]	[-0.091, 0.048]
N	19407	26999	19407	26999	19407	26999
RD effect bandwidth	0.53	0.74	0.53	0.74	0.53	0.74
Bias bandwidth	0.53	0.74	0.53	0.74	0.53	0.74
Bandwidth implementation	IK	CCT	IK	CCT	IK	CCT
Panel C: CCT-correction, bias and RD effect bandwidths optimized for each polynomial specification						
	(13)	(14)	(15)	(16)	(17)	(18)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.030	0.046	0.026	0.052	0.025	0.052
95% confidence interval (non-clustered)	[-0.001, 0.060]	[0.022, 0.069]	[-0.007, 0.060]	[0.030, 0.073]	[-0.012, 0.062]	[0.031, 0.072]
95% confidence interval (clustered)	[-0.004, 0.063]	[0.020, 0.071]	[-0.009, 0.061]	[0.030, 0.073]	[-0.014, 0.064]	[0.031, 0.072]
N	19407	26999	54464	78469	70576	112398
RD effect bandwidth	0.53	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 implementation	IK	CCT	IK	CCT	IK	CCT

Notes: 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 municipality level. 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.

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 bandwidth, but for different RD effect bandwiths. For this figure, we use the IK implementation to determine the bias bandwidth, because it seemed to lead to narrower bandwidths and worked well.

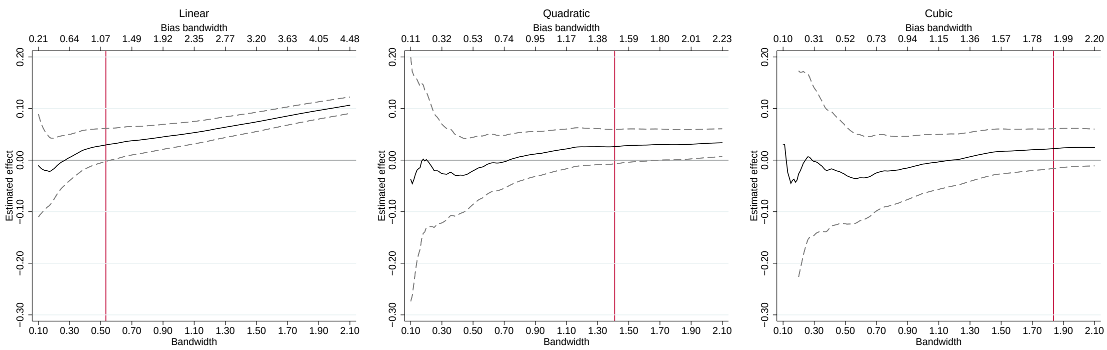
The figure shows that when fixing the bias bandwidth to be IK optimal, the estimated effect is quite robust to the choice of the RD effect bandwidth and most of the time not significantly different from zero.



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. Red 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.

Figure 3. Bias-corrected RDD estimates, fixed bias bandwidth.

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 et al. (2016a) argue that a bandwidth adjustment ("shrinkage") is called for to achieve better coverage error rates when MSE-optimal bandwidths are used.



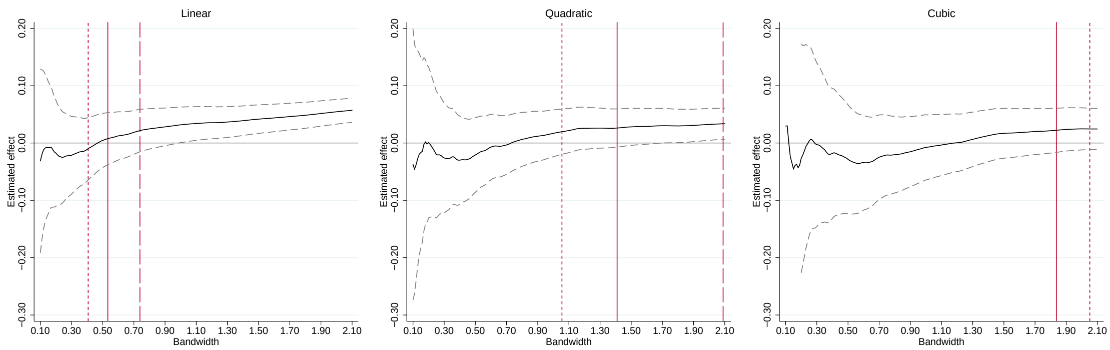
Notes: Figure displays bias-corrected point estimates from local polynomial regressions with triangular kernel using various bandwidths. Dashed lines show 95 % confidence intervals computed using robust standard errors. In the third graph, confidence intervals are omitted for bandwidths smaller than 0.2. Red vertical lines mark the IK RD effect and bias bandwidths (both for estimation and bias correction).

Figure 4. Bias-corrected RDD results, both bandwidths vary.

Furthermore, following the recommendation of Calonico et al. (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* under-smoothing) 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 et al. (2014a) and Calonico et al. (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 et al. (2016a). Unlike *ad hoc* under-smooting, the adjustment improves the coverage error rates of the MSE-optimal bandwidths. For our sample size, the proposed adjustment factors are be 0.55, 0.51 and 0.51 for the linear, quadratic and cubic specification, respectively. As can be seen from Figure 5, applying the shrinkage 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]

[32] To 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% CI of [0.009, 0.060]).



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. In the third graph, confidence intervals are omitted for bandwidths smaller than 0.2. Red vertical solid lines mark the IK bandwidth, long-dashed the CCT bandwidth and short-dashed the adjusted CCT bandwidth. For the cubic specification, the non-adjusted CCT bandwidth does not fit within the x-axis.

Figure 5. Bias-corrected RDD results, bandwidths equal.

We believe that the above results are useful and of interest to applied econometricians, because most of the existing published and on-going work applying RDD uses the same implementations of the MSE-optimal bandwidth and bias-correction approach as 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 et al. 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 invalidate the use of RDD also in the closest races. We find some evidence that there are fewer rejections of covariate balance when more flexible local polynomial specifications (or under-smoothing) 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 under-smoothing 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 too inflexible specifications.

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. Taking 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 (1^{st}–5^{th} degree). These parametric RDD generates 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 the one 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 suggests no incumbency advantage when this alternative measure is used. In contrast, the RDD results suggest a positive effect when RDD is implemented in a standard fashion, using the local linear polynomial and various (MSE) optimal bandwidths.

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 both in 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

[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.

the CCT and IK implementations. The updated software allows for clustering when calculating standard errors and bandwidths. We have re-estimated the most relevant specifications of the previous sections using the new implementation of the 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 et al. 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 et al. (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

[34] Moreover, 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).

[35] As Cattaneo et al. (2014), Cattaneo et al. (2016b), 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.

knowledge, the experiment is unique in the literature, because it takes place exactly at the cutoff. This means that both the experiment and RDD target the same treatment effect.[36]

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 zero 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 at 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 et al. (2014a) is able to recover the experimental benchmark, provided not too wide bias bandwidths are used. We find that the procedure is less sensitive to the choice of the bandwidth (than *ad hoc* under-smoothing) 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 under-smoothing, the method of Calonico et al. (2014a) is more efficient and has faster vanishing coverage error rates. Our findings corroborate the findings of the simulation and formal analyses of Calonico et al. (2014a) and Calonico et al. (2016a), which demonstrate that the method of Calonico et al. (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.

[36] To 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 zero. As Cattaneo et al. (2016) 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.

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* (Advances in Econometrics Vol. 38) M. D. Cattaneo and J. C. Escanciano (ed.), pp. 383-420, Emerald Publishing Limited.

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.

Broockman, D. E. (2009), "Do congressional candidates have reverse coattails? Evidence from a regression discontinuity design." *Political Analysis* 17 (4), 418-434.

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.

Calonico, S., M. D. Cattaneo, and M. F. Farrell (2016a), "On the effect of bias estimation on coverage accuracy in non-parametric inference." Forthcoming in *Journal of the American Statistical Association*.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a), "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica* 82 (6), 2295-2326.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014b), "Robust data-driven inference in the regression discontinuity design." *Stata Journal* 14 (4), 909-946.

Calonico, S., M. D. Cattaneo, M. F. Farrell, and R. Titiunik (2016b), "rdrobust: software for regressions discontinuity designs." *Stata Journal* 17 (2): 372-404.

Card, D., D. S. Lee, Z. Pei, and A. Weber (2014), "Local polynomial order in regressions discontinuity designs." Brandeis University Working Paper 81.

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.

Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare (2016a), "Inference in regression discontinuity designs under local randomization." *Stata Journal* 16 (2), 331-367.

Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare (2016b), "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.

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.

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.

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.

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.

Dahlgaard, J. O. (2016), "You just made it: Individual incumbency advantage under proportional representation." *Electoral Studies* 44, 319-328.

Dehejia, R. and S. Wahba (2002), "Propensity score-matching methods for nonexperimental causal studies." *Review of Economics and Statistics*, 84 (1), 151-161.

De Magalhaes, L. (2014), "Incumbency effects in a comparative perspective: evidence from Brazilian mayoral elections." *Political Analysis* 23 (1), 113-126.

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.

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.

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.

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.

Erikson, R. S. and K. Rader (2017), "Much ado about nothing: rdd and the incumbency advantage." *Political Analysis* 25 (2), 269-275.

Fan, J. and I. Gijbels (1996), "*Local Polynomial Modeling and its Applications*." New York, Chapham & Hall / CRC Press.

Ferreira, F. and J. Gyourko (2009), "Do political parties matter? evidence from U.S. cities." *Quarterly Journal of Economics* 124 (1), 399-422.

Folke, O. and J. M. Snyder (2012), "Gubernatorial midterm slumps." *American Journal of Political Science* 56 (4), 931-948.

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

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.

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.

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.

Gelman, A. and G. King (1990), "Estimating incumbency advantage without bias." *American Journal of Political Science* 34 (4), 1142-1164.

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.

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.

Grimmer, J., E. Hersh, B. Feinstein, and D. Carpenter (2012), "Are close elections random?" Working paper.

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.

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.

Iltasanomat. 4/12/2011. *Tiesitkö tätä vaaleista? Näin käy, jos äänet menevät tasan*. http://www.iltasanomat.fi/vaalit2011/art-1288381910312.html.

Imbens, G. and T. Lemieux (2008), "Regression discontinuity designs: a guide to practice." *Journal of Econometrics* 142 (2), 615-635.

Imbens, G. and K. Kalyanaraman (2012), “Optimal bandwidth choice for the regression discontinuity estimator.” *Review of Economic Studies* 79 (3), 933-959.
Klašnja, M. and R. Titiunik (2016), “The incumbency curse: weak parties, term limits, and unfulfilled accountability.” *American Political Science Review* 111 (1), 129-148.
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.
LaLonde, R. J. (1986), “Evaluating the econometric evaluations of training programs with experimental data.” *American Economic Review* 76 (4), 604-620.
Lee, D. S. (2008), “Randomized experiments from non-random selection in U.S. House elections.” *Journal of Econometrics* 142 (2), 675-697.
Lee, D. S., and T. Lemieux (2010), “Regression discontinuity designs in economics.” *Journal of Economic Literature* 48 (2), 281-355.
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.
Lemieux, T. and K. Milligan (2008), “Incentive effects of social assistance: A regression discontinuity approach.” *Journal of Econometrics* 142, 807–828.
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.
Lundqvist, H. (2011), “Is it worth it? on the returns to holding political office.” Working paper.
McCrary, J. (2008), “Manipulation of the running variable in the regression discontinuity design: a density test.” *Journal of Econometrics* 142 (2), 698-714.
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.
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.
Skovron, C. and R. Titiunik (2015), “A practical guide to regression discontinuity designs in political science.” Working Paper.
Smith, J. and P. E. Todd. (2005), “Does matching overcome LaLonde’s critique of nonexperimental estimators?” *Journal of Econometrics* 125, 305-353.
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.
Snyder, J. (2005), “Detecting manipulation in U.S. House elections.” Working paper.
Stone, P. (2011), “*The Luck of Draw: The Role of Lotteries in Decision-Making*.” Oxford University Press.
The Atlantic. 11/19/2012. When a state election can be literally determined by a coin toss. http://www.theatlantic.com/politics/archive/2012/11/when-a-state-election-can-be-literally-determined-by-a-coin-toss/265413/.
The Telegraph India. 2/7/2014. In election tie, pick winner by lottery. http://www.telegraphindia.com/1140207/jsp/nation/story_17909869.jsp#.VCUt8hCfjjV.
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.

Time. 5/17/2013. Coin toss determines mayor in Philippine town. http://newsfeed.time.com/2013/05/17/coin-toss-determines-mayor-in-philippine-town/.

Trounstine, J. (2011), "Evidence of a local incumbency advantage." *Legislative Studies Quarterly* 36(2), 255-280.

Uppal, Y. (2009), "The disadvantaged incumbents: estimating incumbency effects in Indian state legislatures." *Public Choice* 138 (1), 9-27.

UPI. 7/14/2014. New Mexico judicial election decided by coin toss. http://www.upi.com/Top_News/US/2014/07/14/New-Mexico-judicial-election-decided-by-coin-toss/6161405361068/.

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.

When Does Regression Discontinuity Design Work? Evidence from Random Election Outcomes

Ari Hyytinen, Jaakko Meriläinen, Tuukka Saarimaa, Otto Toivanen and Janne Tukiainen

August 23, 2017

ONLINE APPENDICES

(Supplementary material for online publication only)

We report here the additional empirical analyses to which the main text refers. The supplement consists of Appendices A–F. Appendix A reports summary statistics for our data. In Appendix B, we describe a number of empirical results for the lottery sample. Appendix C characterizes graphically the forcing variable used in the regression discontinuity design (RDD). In Appendix D, we evaluate the validity of the RDD. A large battery of robustness checks is reported in Appendix F. Appendix E reports covariate balance tests for various RDD samples, determined by different bandwidth choices, as well as a brief evaluation of the local randomization assumption.

Appendix A: Supplementary information to Section 2.2 (Data)

In this appendix, we report summary statistics for our data.

Table A1: This table reports descriptive statistics for the individual candidates. As the table shows, the variables that can be regarded as (rough) measures of candidate quality: Many of them obtain, on average, higher values for the elected candidates. For example, the elected candidates have higher income, are more often university-educated and are less often unemployed. The difference is particularly striking when we look at incumbency status: 58% of the elected candidates were incumbents, whereas only 6% of those who were not elected were incumbents.

Table A1. Descriptive statistics for individual candidates.

	All data (N = 198118)			Elected (N = 56734)			Not elected (N = 141384)		
Variable	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Elected next election (only re-runners)	82946	0.38	0.48	32070	0.79	0.41	50876	0.12	0.32
Elected next election (all candidates)	160727	0.19	0.40	46982	0.54	0.50	113745	0.05	0.22
Running next election	160727	0.52	0.50	46982	0.68	0.47	113745	0.45	0.50
Number of votes next election	82946	76	180	32070	131	268	50876	41	65
Vote share next election	82946	1.14	1.31	32070	2.05	1.54	50876	0.57	0.68
Vote share	198117	0.97	1.20	56734	2.22	1.50	141383	0.46	0.47
Number of votes	198117	61	149	56734	127	257	141383	34	45
Female	198118	0.39	0.49	56734	0.35	0.48	141384	0.40	0.49
Age	198117	46.75	12.64	56734	48.15	11.15	141383	46.18	13.15
Incumbent	198118	0.21	0.41	56734	0.58	0.49	141384	0.06	0.24
Municipal employee	160993	0.23	0.42	47060	0.27	0.44	113933	0.22	0.41
Wage income	117787	23738	26978	34566	27813	41548	83221	22045	17417
Capital income	117787	2650	35446	34566	4775	61116	83221	1767	14973
High professional	198022	0.19	0.40	56721	0.24	0.43	141301	0.18	0.38
Entrepreneur	198022	0.15	0.36	56721	0.23	0.42	141301	0.12	0.33
Student	198022	0.04	0.20	56721	0.02	0.13	141301	0.05	0.22
Unemployed	198022	0.07	0.25	56721	0.03	0.18	141301	0.08	0.27
University degree	159437	0.16	0.37	46711	0.20	0.40	112726	0.14	0.35
Coalition Party	198118	0.15	0.36	56734	0.15	0.35	141384	0.16	0.36
Social Democrats	198118	0.18	0.38	56734	0.18	0.38	141384	0.18	0.38
Center Party	198118	0.22	0.42	56734	0.30	0.46	141384	0.19	0.40
True Finns	198118	0.02	0.15	56734	0.01	0.12	141384	0.03	0.16
Green Party	198118	0.04	0.19	56734	0.02	0.15	141384	0.04	0.20
Socialist Party	198118	0.09	0.29	56734	0.07	0.26	141384	0.10	0.30
Swedish Party	198118	0.03	0.17	56734	0.04	0.20	141384	0.02	0.16
Christian Party	198118	0.04	0.18	56734	0.03	0.16	141384	0.04	0.19
Other parties	198118	0.23	0.42	56734	0.20	0.40	141384	0.24	0.43

Notes: Income data are not available for 2012 elections, and in 1996 elections they are available only for candidates who run also in 2000, 2004 and 2008 elections. Income is expressed in euros. Municipal employee status is not available for 2012 elections.

Table A2: This table reports descriptive statistics for municipalities, measured using the candidate level data. As can be seen (the panel on the left), there are three major parties in Finland. The three largest parties' seat shares total to over 70%. There are two main reasons why there are differences in the variables related to elections between the elected candidates' municipalities (the panel in the middle) and the not-elected candidate's municipalities (the panel on the right). First, a larger share of all running candidates is elected in smaller municipalities. For example, the Center Party has a larger vote share in smaller municipalities. Second, there are more candidates in the larger municipalities. The table also shows that in a number of dimensions, like income, age and unemployment rate, there are no major differences in the municipal characteristics between elected and non-elected candidates.

Table A2. Descriptive statistics for municipalities.

Municipality characteristics

	All data (N = 198118)			Elected (N = 56734)			Not elected (N = 141384)		
Variable	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Total number of votes	198118	19935	43682	56734	10607	26431	141384	23677	48421
Coalition Party seat share	198118	19.58	10.10	56734	17.61	10.52	141384	20.38	9.81
Social Democrats seat share	198118	21.88	10.21	56734	20.62	10.88	141384	22.38	9.88
Center Party seat share	198118	30.58	20.52	56734	35.20	21.14	141384	28.73	19.97
True Finns seat share	198118	3.77	5.87	56734	3.49	5.87	141384	3.88	5.86
Green Party seat share	198118	4.25	5.41	56734	2.89	4.30	141384	4.79	5.70
Socialist Party seat share	198118	8.57	7.37	56734	8.14	7.72	141384	8.74	7.22
Swedish Party seat share	198118	4.39	13.87	56734	5.19	16.80	141384	4.07	12.49
Christian Party seat share	198118	3.41	3.56	56734	3.24	3.79	141384	3.48	3.47
Other parties' seat share	198118	3.45	6.74	56734	3.50	7.56	141384	3.43	6.39
Voter turnout	196329	62.20	6.28	56174	63.40	6.28	140155	61.72	6.21
Population	197307	43407	95692	56581	22944	58177	140726	51634	106027
Share of 0-14-year-olds	196385	17.84	3.28	56331	17.96	3.47	140054	17.79	3.20
Share of 15-64-year-olds	196385	64.41	3.48	56331	63.49	3.27	140054	64.78	3.49
Share of over-65-year-olds	196385	17.75	4.82	56331	18.55	4.99	140054	17.43	4.72
Income per capita	196385	21204	5876	56331	20364	5634	140054	21543	5937
Unemployment	197307	13.50	5.71	56581	13.77	5.85	140726	13.39	5.65

Notes: Income per capita is expressed in euros.

Appendix B: Supplementary information to Section 3.1 (Experimental estimates)

In this appendix, we report a number of empirical results obtained using the lottery sample (i.e., the sample which only includes the candidates that had a tie). These results bear on the robustness of the experimental estimate.

Table B1: This table shows additional balance checks for party affiliation and municipality characteristics in the lottery sample. These characteristics should be balanced by construction, as we construct the forcing variable within party lists. The table shows that the samples are, indeed, almost identical. The small and insignificant differences in the means are likely due to the fact that in some lotteries there are more than two candidates.

Table B1. Additional balance checks.

Individual characteristics

Variable	Elected (N = 671) N	Mean	Std. Dev.	Not elected (N = 680) N	Mean	Std. Dev.	Difference
Coalition Party	671	0.20	0.40	680	0.20	0.40	0.00
Social Democrats	671	0.18	0.39	680	0.18	0.39	0.00
Center Party	671	0.42	0.49	680	0.42	0.49	0.00
True Finns	671	0.02	0.13	680	0.02	0.13	0.00
Green Party	671	0.01	0.11	680	0.01	0.11	0.00
Socialist Party	671	0.08	0.27	680	0.08	0.27	0.00
Swedish Party	671	0.03	0.18	680	0.04	0.19	-0.01
Christian Party	671	0.02	0.15	680	0.02	0.15	0.00
Other parties	671	0.03	0.18	680	0.03	0.18	0.00

Municipality characteristics

Variable	Elected (N = 671) N	Mean	Std. Dev.	Not elected (N = 680) N	Mean	Std. Dev.	Difference
Total number of votes	671	4467	12006	680	4395	11921	71
Coalition Party seat share	671	16.88	11.08	680	16.76	10.88	0.13
Social Democrats seat share	671	19.70	10.76	680	19.63	10.95	0.07
Center Party seat share	671	41.46	19.98	680	41.57	20.17	-0.11
True Finns seat share	671	1.92	4.79	680	1.89	4.59	0.02
Green Party seat share	671	1.72	3.29	680	1.73	3.31	-0.01
Socialist Party seat share	671	7.55	7.91	680	7.56	7.82	0.00
Swedish Party seat share	671	3.70	14.42	680	3.97	14.95	-0.27
Christian Party seat share	671	2.87	3.92	680	2.83	3.92	0.04
Other parties' seat share	671	3.76	8.59	680	3.63	8.48	0.13
Voter turnout	664	65.23	5.90	673	65.38	6.02	-0.15
Population	671	9316	25430	680	9145	25241	171
Share of 0-14-year-olds	667	18.31	3.31	676	18.42	3.33	-0.11
Share of 15-64-year-olds	667	62.97	2.87	676	62.89	2.90	0.07
Share of over-65-year-olds	667	18.72	4.69	676	18.69	4.68	0.03
Income per capita	667	18457	5372	676	18413	5372	44
Unemployment	671	14.85	6.75	680	14.80	6.69	0.05

Notes: Differences in means have been tested using t test adjusted for clustering at municipality level. Sample includes only candidates running in 1996-2008 elections. Income data are not available for 2012 elections, and in 1996 elections they are available only for candidates who run also in 2000, 2004 and 2008 elections. Income and income per capita are expressed in euros.

Table B2: This table reports experimental results for the alternative outcomes, vote share (Panel A) and running (Panel B) in the next elections. The regressions use the entire lottery sample. They provide no evidence of personal incumbency advantage. We have also checked that the effect is close to zero and not significant if the absolute number of votes in the next election is used as the outcome variable (not reported).

Table B2. Experimental results for alternative outcomes.

Panel A: Vote share next election				
	(1)	(2)	(3)	(4)
Elected	0.012	0.006	-0.020	-0.014
95% confidence interval	[-0.102, 0.125]	[-0.108, 0.121]	[-0.152, 0.111]	[-0.160, 0.133]
N	1351	1351	1351	1351
R^2	0.00	0.06	0.37	0.52
Panel B: Running next election				
	(5)	(6)	(7)	(8)
Elected	0.011	0.007	0.001	0.005
95% confidence interval	[-0.040, 0.062]	[-0.044, 0.058]	[-0.058, 0.059]	[-0.060, 0.071]
N	1351	1351	1351	1351
R^2	0.00	0.05	0.30	0.45
Controls	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	No
Municipality-year fixed ef	No	No	No	Yes

Notes : Only actual lotteries are included in the regressions. Vote share is set to zero for those candidates that do not run in the next election. Set of controls includes age, gender, party affiliation, socio-economic status and incumbency status of a candidate, and total number of votes. Some specifications include also municipality or municipality-year fixed effects.Confidence intervals are based on standard errors clustered at the municipality level. Unit of observation is a candidate *i* at year *t* .

Table B3: In this table, we look at elections in small and large municipalities separately. We split the sample based on the median number of total votes in the municipality in the lottery sample. This median is 2422. The median is slightly higher (2662) in the entire sample. The regressions reported in the table below do not include any controls. They should therefore be compared to the result in column (1) in Table 2 in the main text of HMSTT. As can be seen from the table, we do not find evidence for an incumbency advantage in either sub-sample.

Table B3. Experimental results for small and large elections.

Outcome: Elected next election

	(1)	(2)
Elected	0.002	0.006
95% confidence interval	[-0.064, 0.067]	[-0.065, 0.077]
N	687	664
R^2	0.00	0.00
Sample	Small elections	Large elections

Notes: An election is considered small (large), if at most (more than) 2422 votes are cast. Only actual lotteries are included in the regressions.Confidence intervals are based on standard errors clustered at municipality level. Unit of observation is a candidate *i* at year *t* .

Table B4: We have reproduced the experimental estimate using a sample from which those who do not rerun are excluded. We report these results for our main outcome and the alternative outcome (the vote share). These results provide no evidence of a personal incumbency advantage.

Table B4. Experimental estimates for rerunners.

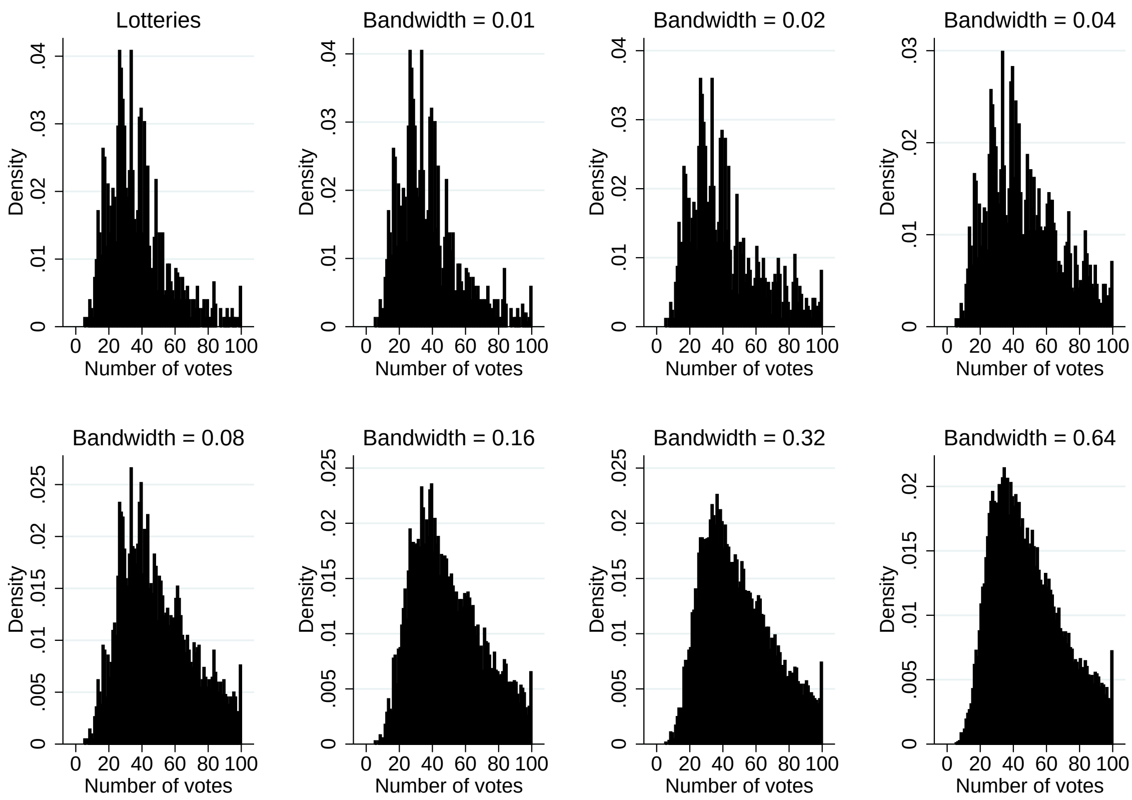
	(1)	(2)	(3)	(4)
Outcome: Elected next election				
Elected	-0.003	-0.002	0.025	0.035
	[-0.071, 0.066]	[-0.073, 0.068]	[-0.073, 0.124]	[-0.091, 0.160]
N	820	820	820	820
R^2	0.00	0.04	0.41	0.64
Outcome: Vote share next election				
	(5)	(6)	(7)	(8)
Elected	-0.012	-0.009	0.051	0.021
	[-0.145, 0.122]	[-0.142, 0.124]	[-0.110, 0.212]	[-0.184, 0.226]
N	820	820	820	820
R^2	0.00	0.17	0.67	0.80
Controls	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	No
Municipality-year fixed effects	No	No	No	Yes

Notes: Only actual lotteries and rerunning candidates are included in the regressions. Set of controls includes age, gender, party affiliation, socio-economic status and incumbency status of a candidate, and total number of votes. Some specifications include also municipality or municipality-year fixed effects. Unit of observation is a candidate *i* at year *t*.

Appendix C: Supplementary information to Section 3.2 (Non-experimental estimates)

This appendix provides additional figures to characterize our forcing variable, v_{it}. We call our forcing variable "Vote margin (%)" in some of the graphs below, where the margin refers to the distance to the cutoff. The forcing variable is reported in percentage points. For example, a value 0.5 refers to 5 votes out of 1000.

Figure C1: In this figure, we graph the distribution of the number of votes within different bandwidths in the forcing variables. The figures show how many votes the candidates involved in close elections receive. The distribution gets a large amount of mass around 30–50 votes.



Notes: Figure shows the distribution of number of votes within one bandwidth on both sides of the cutoff for different bandwidths. Bin size is 1 vote. x-axis is restricted to 100 votes.

Figure C1. The distribution of the number of votes for different bandwidths.

Figure C2: This figure displays the relationship between the forcing variable and the distance to cutoff (vote distance), as measured by the absolute number of votes. The density graphs show that, as expected, the candidates are further away from the cutoff in terms of absolute number of votes as the bandwidth becomes wider. For all reported bandwidths, the most common distance is only one or two votes.

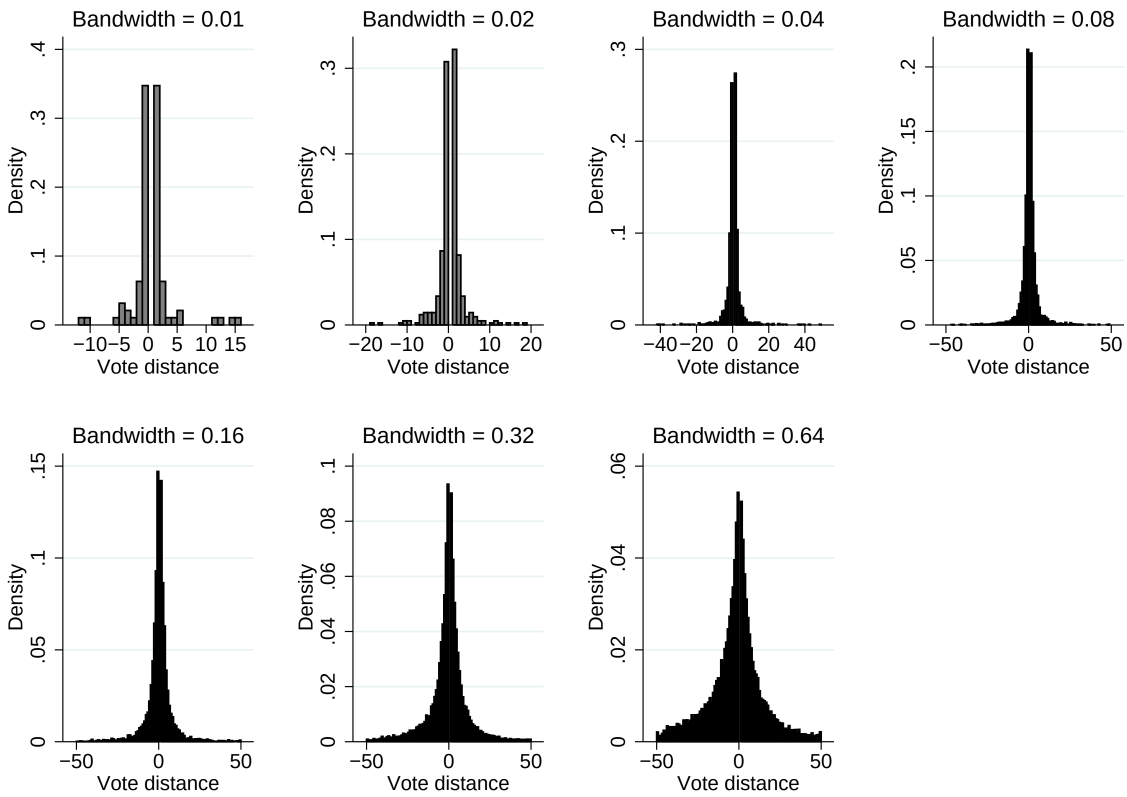


Figure C2. Distribution of the distance to cutoff in absolute votes for different bandwidths of the forcing variable.

Figure C3: This figure maps the relationship between the forcing variable (vote margin, x-axis) and the distance to cutoff measured in the absolute number of votes (y-axis). It shows that, overall, the two are positively correlated within the reported bandwidth. There are fairly many observations also on or close by the horizontal line. This means that, within the reported bandwidth, for each value of the forcing variable there are many observations that are only one or two votes from the cutoff. This echoes what Figure C2 shows.

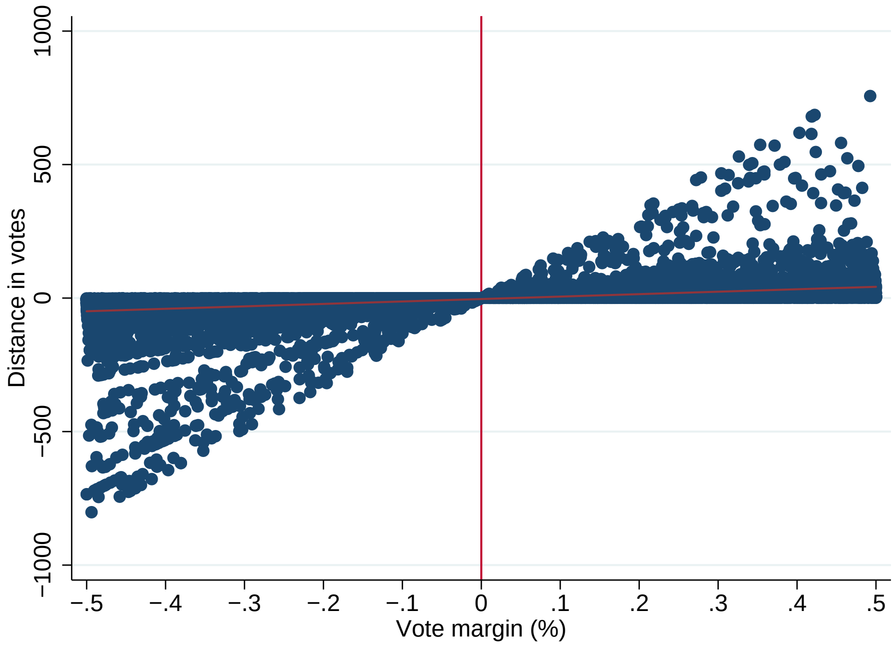
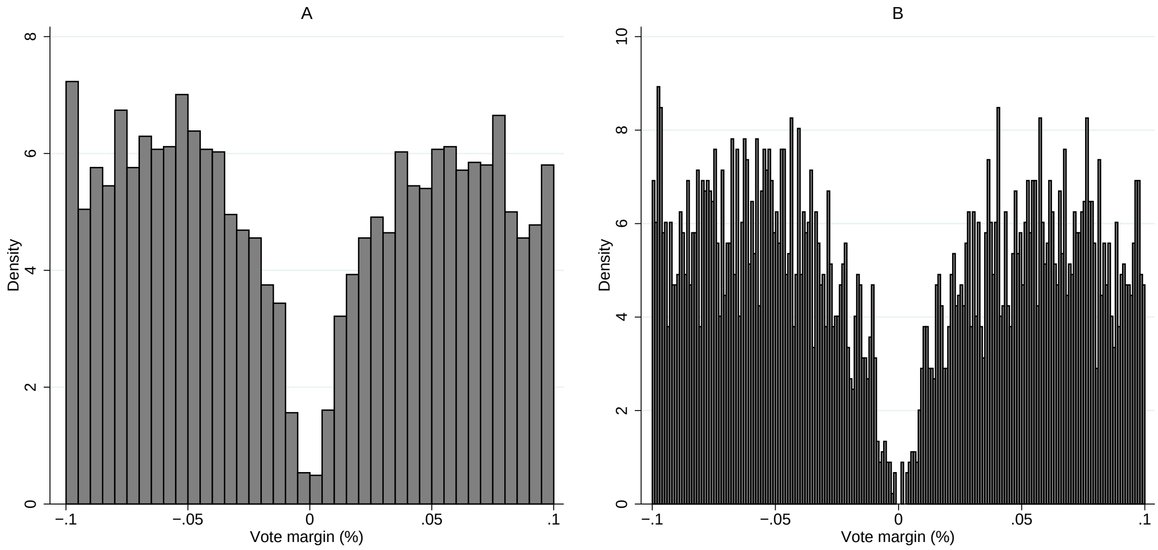


Figure C3. Relationship between the forcing variable and the distance to cutoff measured in absolute votes.

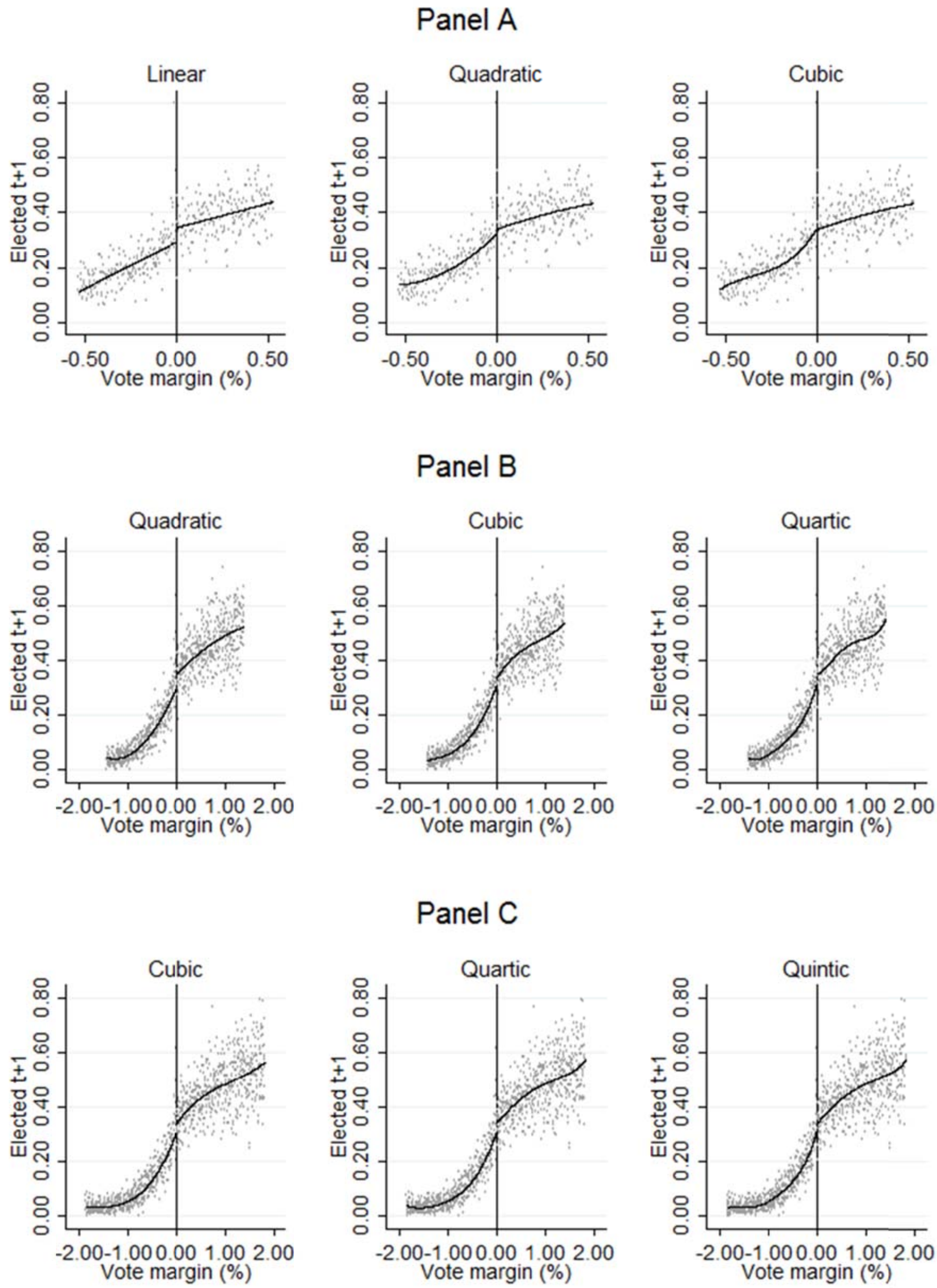
Figure C4: These histograms show the distribution of the forcing variable within two very small bandwidths nearby the RDD cutoff. The histograms suggest that the forcing variable can be treated as continuous for the purposes of RDD. The dip in the density of the forcing variable between -0.01 and 0.01 is related to the fact that the forcing variable can obtain such small values only when the party lists are large. For example, a value of 0.01 refers to one vote out of ten thousand. Lists that get more than ten thousand votes exist only in the larger municipalities.



Notes: Figure A shows histogram of the forcing variable with bins of 0.005, and figure B uses bins of 0.001. Values of the forcing variable are limited between -0.1 and 0.1. Lotteries have been excluded.

Figure C4. Histogram of the forcing variable close to the cutoff.

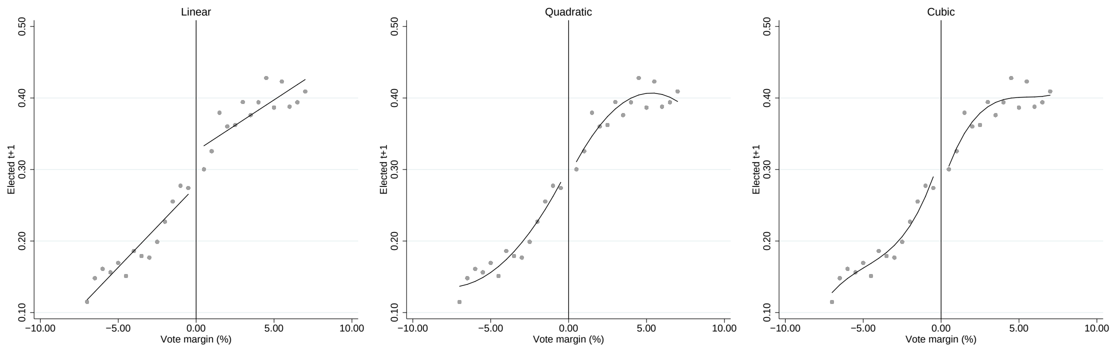
Figure C5: These figures are similar to Figure 1 in the main text, but they give a richer picture of the underlying data, as they show the binned averages within a larger number of bins. These bins have been chosen applying mimicking variance evenly spaced method using spacing estimators (see Calonico et al. 2015). We estimate the optimal Imbens-Kalyanaraman bandwidth for the left-most specification in each panel, and then increase the degree of the control polynomial by one or two.



Notes: Figure shows local polynomial fits with a triangular kernel within the optimal Imbens-Kalyanaraman (2012) bandwidth optimized for the linear specification in Panel A, quadratic specification in Panel B and 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 the optimal bandwidths are calculated for. In the midmost 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 chosen using mimicking variance evenly-spaced method using spacing estimators (see Calonico et al. 2015).

Figure C5. Curvature between the forcing variable and the outcome

Figure C6: These figures display RDD fit and a scatter of plot of observation bins around the cutoff when the forcing variable is defined as the (non-normalized) number of votes. The main purpose of these figures is to show that the documented features in the relationship between the forcing variable and outcome are not unique to the way we define the forcing variable in the main text. This indeed appears not to be the case: As the figures show, there is a clear jump at the cutoff in the figure on the left and evidence of curvature in the middle and on the right.



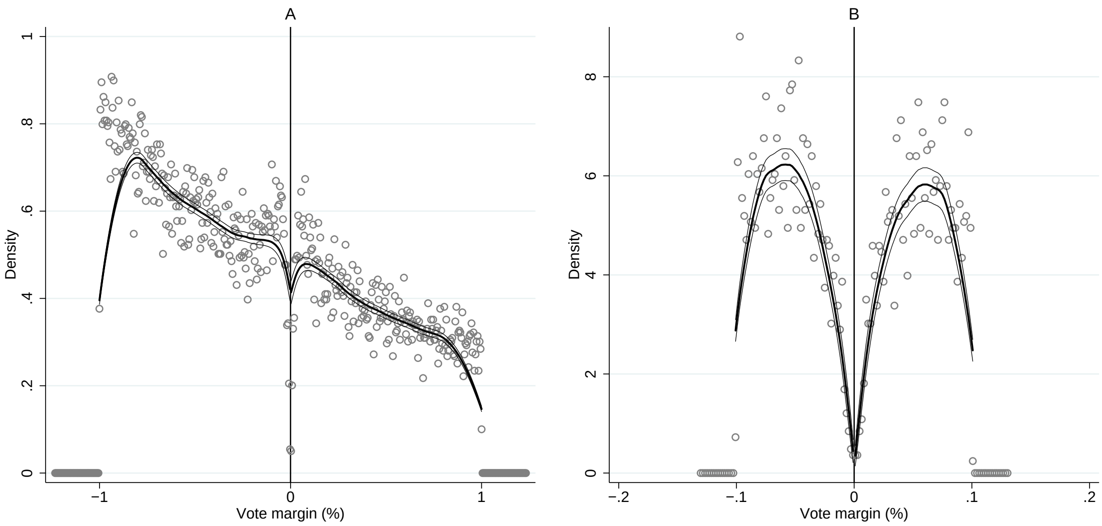
Notes: Figure shows local polynomial fits with triangular kernel within the optimal Imbens-Kalyanaraman (2012) bandwidth optimized for the linear specification. Gray dots mark binned averages.

Figure C6. Curvature between the non-scaled forcing variable (number of votes) and the outcome

Appendix D: Supplementary information to HMSTT Section 4.1 (RDD falsification and smoothness tests)

In this appendix, we report validity tests to for RDD. The reported pattern of validity tests includes i) the McCrary (2008) manipulation test, ii) covariate balance tests, and iii) placebo tests where the location of the cutoff is artificially redefined.

Figure D1: This figure reports the McCrary (2008) tests. The test asks whether there is a jump in the amount of observations at the cutoff of getting elected. Such jump would indicate that some candidates have been able to manipulate into getting the treatment. There is no jump. The estimated difference in height is -0.0140 (standard error 0.0474) in graph A (the values of the forcing variable restricted between -1 and 1), and -0.5701 (standard error 0.6616) in graph B (the values of the forcing variable restricted between -0.1 and 0.1). This is not surprising, since there cannot be a jump in the amount of candidates elected: The number of council seats available is fixed. If one candidate is able to manipulate into getting elected, another candidate will not be elected.



Notes: Graph A shows the McCrary (2008) density test with the forcing variable within -1 and 1. Graph B shows the density test with forcing variable within -0.1 and 0.1.

Figure D1. McCrary density test.

Table D1: The main identification assumption in RDD is that covariates develop smoothly over the cutoff. The recent literature (e.g. Snyder et al. 2015 and Eggers et al. 2015) argues that especially in close election applications, balance tests based on the comparisons of means across the cutoff are likely to (wrongly) signal imbalance, because the covariates may vary strongly with the forcing variable near the cutoff. One should, therefore, control for this co-variation ("slopes") when implementing the balance tests. Panel A of Table D1 uses therefore the optimal bandwidth for the local linear specification computed for each covariate separately. When testing for covariate smoothness, bandwidth needs to be optimized for each covariate separately, because they are each unique in their relation to the forcing variable. We report in Panel B of Table D1 also the results that use half the optimal bandwidth. We do so to check how under-smoothing influences the covariance balance tests and to make sure that curvature issues (similar to those we report for our main outcome) do not lead to wrong conclusions about the covariate balance. If some of the covariates have a lot of curvature nearby the cutoff, one might wrongly infer that there is imbalance unless under-smoothing, or some other de-biasing method, is used to obtain more valid confidence intervals.

As can be seen from Panel A and B, there are some significant estimates. We cannot rule out that the few imbalances are due to multiple testing, because Panel A and B are not completely in line with each other in this regard. It is also possible that the estimated jumps are due to substantial curvature in the relationship between the given covariate and the forcing variable near the cutoff. This seems to be at least partly the case, since many of the jumps are no longer statistically significant when more flexible specifications (smaller bandwidths for a given local polynomial or higher order polynomials for a given bandwidth) are used. This means that there are fewer rejections of covariate balance when more flexible local polynomial specifications (or under-smoothing) are used.

We conclude that, taken together, the covariate balance tests provide somewhat mixed evidence. Overall, they do not cast clear doubt on the validity of RDD.

Table D1. Covariate smoothness test.

Panel A: Bandwidth optimized for local linear specification

	(1)	(2)	(3)		(4)	(5)	(6)
	Linear	Quadratic	Cubic		Linear	Quadratic	Cubic
Vote share	-0.005	-0.009	-0.004	Number of votes	3.060**	2.745	-1.501
	(0.009)	(0.011)	(0.014)		(1.120)	(1.421)	(3.046)
N	27125	27125	27125	N	37155	37155	37155
Bandwidth	0.74	0.74	0.74	Bandwidth	0.99	0.99	0.99
	(7)	(8)	(9)		(10)	(11)	(12)
	Linear	Quadratic	Cubic		Linear	Quadratic	Cubic
Female	-0.037**	-0.018	-0.007	Age	0.349	-0.248	-0.225
	(0.008)	(0.010)	(0.013)		(0.216)	(0.323)	(0.469)
N	94185	94185	94185	N	71196	71196	71196
Bandwidth	2.74	2.74	2.74	Bandwidth	1.85	1.85	1.85
	(13)	(14)	(15)		(16)	(17)	(18)
	Linear	Quadratic	Cubic		Linear	Quadratic	Cubic
Incumbent	0.019	0.022	0.023	Municipal employee	0.000	0.001	0.005
	(0.012)	(0.019)	(0.027)		(0.006)	(0.008)	(0.010)
N	27450	27450	27450	N	107961	107961	107961
Bandwidth	0.75	0.75	0.75	Bandwidth	3.60	3.60	3.60
	(19)	(20)	(21)		(22)	(23)	(24)
	Linear	Quadratic	Cubic		Linear	Quadratic	Cubic
Wage income	91	756	338	Capital income	-225	-248	284
	(460)	(638)	(782)		(634)	(1104)	(1743)
N	39252	39252	39252	N	38572	38572	38572
Bandwidth	1.37	1.37	1.37	Bandwidth	1.35	1.35	1.35
	(25)	(26)	(27)		(28)	(29)	(30)
	Linear	Quadratic	Cubic		Linear	Quadratic	Cubic
High professional	-0.0269**	-0.006	0.010	Entrepreneur	0.016*	0.001	-0.011
	(0.008)	(0.009)	(0.013)		(0.008)	(0.011)	(0.014)
N	93021	93021	93021	N	60119	60119	60119
Bandwidth	2.69	2.69	2.69	Bandwidth	1.55	1.55	1.55
	(31)	(32)	(33)		(34)	(35)	(36)
	Linear	Quadratic	Cubic		Linear	Quadratic	Cubic
Student	-0.005	-0.005	-0.007	Unemployed	0.004	0.006	0.007
	(0.003)	(0.004)	(0.005)		(0.003)	(0.005)	(0.006)
N	77124	77124	77124	N	78963	78963	78963
Bandwidth	2.04	2.04	2.04	Bandwidth	2.10	2.10	2.10
	(37)	(38)	(39)				
	Linear	Quadratic	Cubic				
University	-0.017	0.009	0.029*				
	(0.010)	(0.010)	(0.014)				
N	71647	71647	71647				
Bandwidth	2.45	2.45	2.45				

Panel B: 0.5 * bandwidth optimized for local linear specification

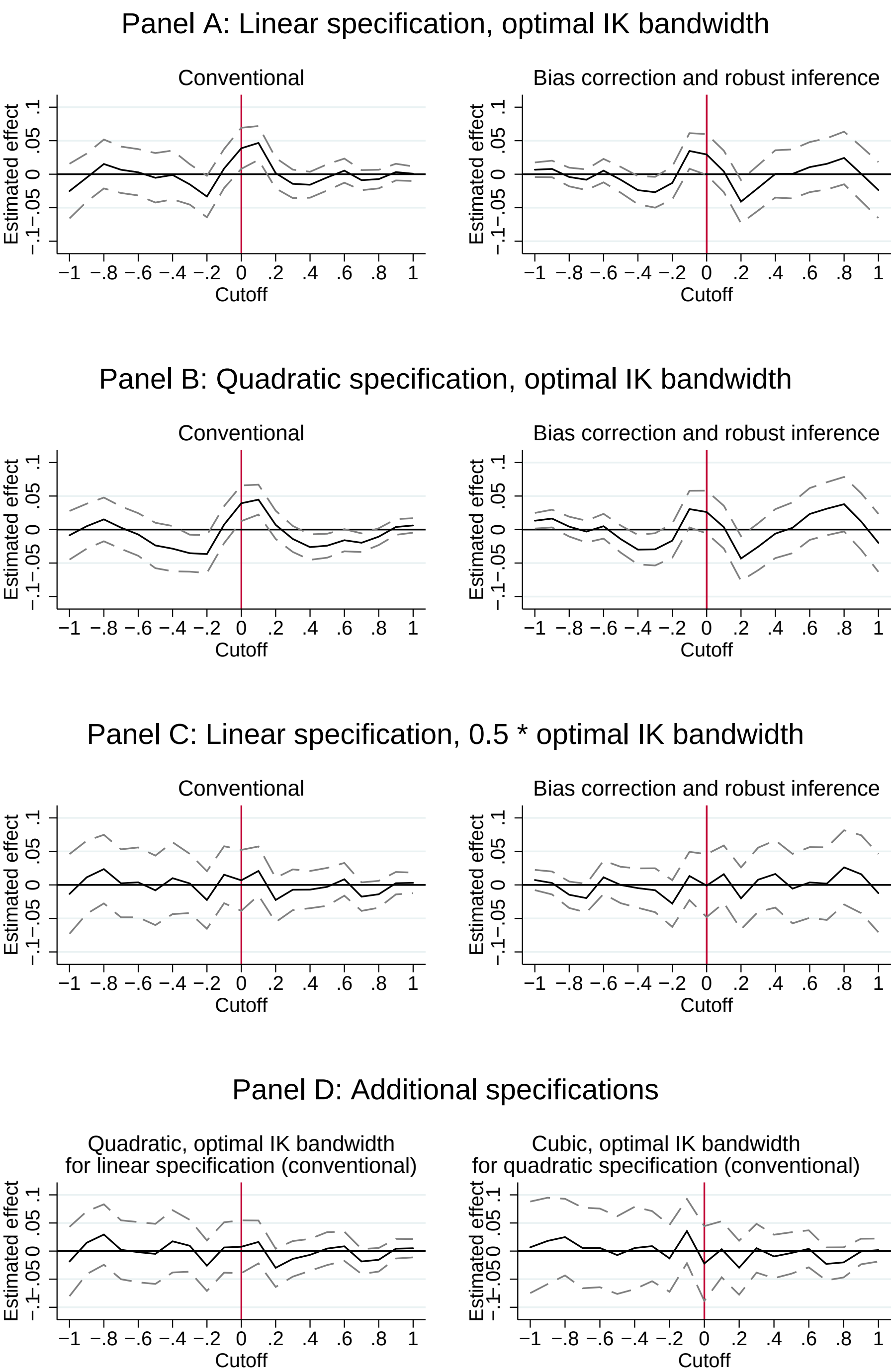
	(37)	(38)	(39)		(40)	(41)	(42)
	Linear	Quadratic	Cubic		Linear	Quadratic	Cubic
Vote share	0.000	0.003	-0.003	Number of votes	0.624	-0.255	2.547
	(0.009)	(0.013)	(0.020)		(1.437)	(1.431)	(5.732)
N	13555	13555	13555	N	18077	18077	18077
Bandwidth	0.37	0.37	0.37	Bandwidth	0.50	0.50	0.50
	(43)	(44)	(45)		(46)	(47)	(48)
	Linear	Quadratic	Cubic		Linear	Quadratic	Cubic
Female	-0.020*	-0.009	-0.014	Age	-0.096	-0.095	-0.272
	(0.010)	(0.014)	(0.019)		(0.316)	(0.475)	(0.651)
N	52883	52883	52883	N	34347	34347	34347
Bandwidth	1.37	1.37	1.37	Bandwidth	0.93	0.93	0.93
	(49)	(50)	(51)		(52)	(53)	(54)
	Linear	Quadratic	Cubic		Linear	Quadratic	Cubic
Incumbent	0.022	0.019	0.011	Municipal employee	0.002	0.007	0.001
	(0.018)	(0.030)	(0.045)		(0.007)	(0.011)	(0.014)
N	13686	13686	13686	N	69385	69385	69385
Bandwidth	0.37	0.37	0.37	Bandwidth	1.80	1.80	1.80
	(55)	(56)	(57)		(58)	(59)	(60)
	Linear	Quadratic	Cubic		Linear	Quadratic	Cubic
Wage income	397	-214	-767	Capital income	-285	642	3321
	(591)	(806)	(1142)		(1106)	(1967)	(2793)
N	18528	18528	18528	N	18205	18205	18205
Bandwidth	0.69	0.69	0.69	Bandwidth	0.67	0.67	0.67
	(61)	(62)	(63)		(64)	(65)	(66)
	Linear	Quadratic	Cubic		Linear	Quadratic	Cubic
High professional	-0.008	0.007	0.010	Entrepreneur	0.001	-0.011	-0.018
	(0.009)	(0.013)	(0.016)		(0.011)	(0.015)	(0.020)
N	51757	51757	51757	N	28379	28379	28379
Bandwidth	1.34	1.34	1.34	Bandwidth	0.77	0.77	0.77
	(67)	(68)	(69)		(70)	(71)	(72)
	Linear	Quadratic	Cubic		Linear	Quadratic	Cubic
Student	-0.005	-0.009	-0.016*	Unemployed	0.006	0.008	0.011
	(0.004)	(0.005)	(0.007)		(0.004)	(0.006)	(0.008)
N	38230	38230	38230	N	39557	39557	39557
Bandwidth	1.02	1.02	1.02	Bandwidth	1.05	1.05	1.05
	(73)	(74)	(75)				
	Linear	Quadratic	Cubic				
University	0.007	0.029*	0.024				
	(0.009)	(0.014)	(0.018)				
N	38403	38403	38403				
Bandwidth	1.22	1.22	1.22				

Notes: Panel A shows estimated discontinuities in covariates using local polynomial regressions within the optimal IK bandwidth for local linear specification. Panel B uses bandwidth half of the optimal. * and ** denote 5% and 1% statistical significance levels, respectively. Unit of observation is a candidate *i* at year *t*.

Figure D2: Figure D2 reports a series of placebo tests where the location of the cutoff is artificially redefined. If there are jumps in locations other than the true cutoff, it would suggest that strong nonlinearities or discontinuities in the relationship between the forcing variable and the outcome may be driving the RDD result (instead of a causal effect at the cutoff). Typically, these tests are used in applications where there is a documented effect at the cutoff (that is statistically different from zero) and the researcher wants to show that this statistically significant jump is unique (or, at least, that only 5% of the placebo cutoffs show jumps that are significant at the 5% level).

In Panel A and B, we display the placebo RDD estimates that are based on the conventional local linear and quadratic specification, using the corresponding IK optimal bandwidths. As we report in the main text, the RDD estimates produced by these specifications indicate that there would be a positive jump at the true cutoff. This is in contrast to what our experimental estimate suggests. As the placebo estimates on the left of these panels show, there also are statistically significant jumps at some of the placebo cutoffs located close by the true cutoff. Some of these jumps are even larger than the one found at the true cutoff. These placebo tests are thus indicative of these RDD specifications not working properly. The placebo graphs on the right have been produced using the same specifications as on the left, but with the CCT-correction. They, too, are indicative of these specifications not working as expected.

In Panels C and D, we explore whether those RDD specifications that in our context seem to work are problematic in the light of the placebo tests. Panel C reports the results for half the optimal (IK) bandwidths: On the left, we use the conventional local linear specification for this under-smoothing approach. The corresponding estimates based on the CCT-correction are displayed on the right. In Panel D we explore whether a polynomial of order $p+1$ is flexible enough for the bandwidth that has been optimized for a polynomial of order p. The panel reports these results for the quadratic and cubic local polynomials. As the two panels show, there are no jumps at any of the placebo cutoffs, implying that these specifications work appropriately. In sum, the placebo tests reported in Panel C and D do suggest that the under-smoothing procedure or the use of higher degree local polynomials without adjusting the bandwidth accordingly may work. These findings thus suggest that the placebo cutoff tests seem to be of use in detecting too inflexible specifications.



Notes: The figure shows the RDD point estimates and the 95% confidence intervals from specifications using local polynomial regression with a triangular kernel. All the left hand graphs and also the right hand graph in Panel D use conventional approach with optimal IK bandwidths and confidence intervals constructed using standard errors clustered by municipality. All the right hand graphs in Panels A-C use IK bandwidth and bias-correction and robust inference by Calonico et al. (2014a). We report the results at various artificial (placebo) cutoffs where the location of the artificial cutoff relative to the true cutoff is reported in the x-axis. In Panel A, bandwidth is optimized for the linear specification, In Panel B, bandwidth is half the one in Panel A and in Panel C, bandwidth is optimized for the quadratic specification. In Panel D, bandwidth is optimized for p-order polynomial specification whereas the fit is based on p+1 order. Optimal bandwidth is based on the specification and sample at the real cutoff. Vertical red line marks the real cutoff.

Figure D2. RDD estimates at the artificial cutoffs.

Appendix E: Supplementary information to Section 4.2 (Robustness tests)

This appendix discusses the robustness tests (#1–#8) that we have conducted.

Robustness test #1: Global polynomial RDD

Table E1: In this table we report results for a parametric RDD specification using higher order *global* polynomials (1^{st}-5^{th} degree) of the forcing variable on both sides of the cutoff. As the table shows, the treatment effect estimates tend to get smaller when the degree of the polynomial increases, but even for the 5^{th} degree polynomial, they are positive, very large in size, and highly significant. The bias using global polynomials seems to an order of magnitude larger than the one obtained using local polynomials. This approach generates incumbency effects that are roughly similar in magnitude to those reported in Lee (2008). It should be noted, however, that his estimates refer to an amalgam of party and personal incumbency effects and apply to a very different institutional context.

Table E1. Parametric RDD with 1^{st}–5^{th} order polynomials.

Outcome: Elected next election

	(1)	(2)	(3)	(4)	(5)
Elected	0.432	0.386	0.342	0.296	0.255
95% confidence interval	[0.422, 0.442]	[0.374, 0.398]	[0.328, 0.355]	[0.281, 0.311]	[0.239, 0.272]
N	154543	154543	154543	154543	154543
R^2	0.33	0.33	0.33	0.34	0.34
Order of control polynomial	1st	2nd	3rd	4th	5th

Notes: Each specification uses the whole range of data. Confidence intervals are based on standard errors clustered at municipality level. Unit of observation is a candidate *i* at year *t*.

Robustness test #2: Alternative measure of incumbency advantage

Table E2: In this table, we look at the effect of being elected in election at time *t* on the vote share in the election at time *t*+1. As we reported earlier (Table B2 in Appendix B), the effect is not statistically different from zero in the lottery sample when this variable is used as an alternative outcome. As the table below shows, the conventional RDD using optimal bandwidths and local linear specification produces a positive and significant effect. The more flexible specifications reproduce the experimental estimate: The estimates suggest that the under-smoothing procedure and the use of higher degree local polynomials without adjusting the bandwidth accordingly work. The bias-correction procedure of Calonico et al. (2014a) reproduces the experimental estimate for this outcome (Panel C). Adjusting the MSE-optimal bandwidths with the adjustment factor suggested by Calonico et al. (2016a) also shows that the RDD estimates are in line with the experimental estimate (Panel D). It is, however, important to point out that some of the estimates in Panel B are negative and quite large in the absolute value.

Table E2. RDD results, incumbency advantage in vote share.

Outcome: Vote share next election						
Panel A: Bandwidth optimized for local linear specification						
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear		Quadratic		Cubic	
Elected	0.049	0.036	0.006	-0.001	-0.019	-0.034
95% confidence interval (clustered)	[0.012, 0.086]	[-0.004, 0.077]	[-0.046, 0.059]	[-0.061, 0.059]	[-0.090, 0.052]	[-0.111, 0.044]
N	36834	28925	36834	28925	36834	28925
Bandwidth	0.99	0.79	0.99	0.79	0.99	0.79
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	0.016	0.007	-0.026	-0.052	-0.086	-0.100
95% confidence interval (clustered)	[-0.034, 0.066]	[-0.048, 0.063]	[-0.100, 0.048]	[-0.136, 0.031]	[-0.187, 0.016]	[-0.213, 0.012]
N	17930	14348	17930	14348	17930	14348
Bandwidth	0.49	0.39	0.49	0.39	0.49	0.39
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT
Panel C: Bandwidths optimized for each specification, CCT-procedure						
	(13)	(14)	(15)	(16)	(17)	(18)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.006	-0.001	-0.003	0.002	-0.015	0.010
95% confidence interval (robust)	[-0.048, 0.060]	[-0.061, 0.058]	[-0.056, 0.050]	[-0.049, 0.053]	[-0.076, 0.046]	[-0.039, 0.058]
N	36834	28925	70205	76855	79078	109826
Bandwidth	0.99	0.79	1.83	2.03	2.11	3.76
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT
Panel D: Adjusted optimal bandwidths for each specification, CCT-procedure						
	(19)	(20)	(21)	(22)	(23)	(24)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	-0.020	-0.042	-0.021	-0.015	-0.045	-0.015
95% confidence interval (robust)	[-0.090, 0.050]	[-0.120, 0.036]	[-0.093, 0.051]	[-0.084, 0.053]	[-0.128, 0.038]	[-0.079, 0.048]
N	19742	15763	34189	38513	40965	73930
Bandwidth	0.54	0.43	0.92	1.03	1.09	1.94
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT

Notes: Table shows estimated incumbency advantage using local polynomial regressions within various bandwidths. Confidence intervals in panels A and B use standard errors clustered at municipality level. Panels C and D use the same main and bias bandwidths. Unit of observation is a candidate *i* at year *t*.

Robustness test #3: Small vs. large municipalities

Tables E3 and E4: These tables reports RDD results separately for small (Table E3) and large (Table E4) municipalities and thus for small and large elections. We use the median number of votes in the municipality in the lottery sample as the point of division (i.e., 2422 votes). As is noted in the main text of HMSTT (and in Appendix B), ties usually appear in elections held in slightly smaller municipalities (those with a small number of voters). This means that our experimental estimate may mostly apply to such elections. As we reported earlier, the experimental estimate is very close to zero both in small and in large elections. However, our forcing variable, v_{it}, can get values really close to zero only when parties get a large amount of votes. This tends to happen in larger elections. The RDD estimates, which use the narrowest bandwidths, may thus mostly apply to them. To check whether the discrepancy between the experimental and the RDD estimates is driven by the size of the municipalities, Tables E3 and E4 reports parts of our RDD analysis separately for small and large municipalities. The results show that our conclusions are not driven by the size of the elections. The bias-correction procedure of Calonico et al. (2014a) reproduces the experimental estimate (Panel C) for IK and CTT bandwidths, except for the cubic specification. Adjusting the MSE-optimal bandwidths with the adjustment factor suggested by Calonico et al. (2016a) brings all the RDD estimates in line with the experimental estimate (Panel D).

Table E3. RDD results for small municipalities.

Outcome: Elected next election						
Panel A: Bandwidth optimized for local linear specification						
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear		Quadratic		Cubic	
Elected (conventional)	0.112	0.036	0.034	0.013	0.011	0.002
95% confidence interval (clustered)	[0.090, 0.135]	[-0.001, 0.072]	[0.001, 0.067]	[-0.044, 0.071]	[-0.033, 0.055]	[-0.076, 0.079]
N	23967	10611	23967	10611	23967	10611
Bandwidth	4.01	1.41	4.01	1.41	4.01	1.41
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 (conventional)	0.051	0.018	0.017	0.007	0.010	0.039
95% confidence interval (clustered)	[0.021, 0.082]	[-0.035, 0.072]	[-0.030, 0.064]	[-0.078, 0.092]	[-0.054, 0.074]	[-0.100, 0.178]
N	14563	5598	14563	5598	14563	5598
Bandwidth	2.00	0.71	2.00	0.71	2.00	0.71
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT
Panel C: Bandwidths optimized for each specification, CCT-procedure						
	(13)	(14)	(15)	(16)	(17)	(18)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.034	0.013	0.014	0.012	0.012	0.010
95% confidence interval (robust)	[0.000, 0.068]	[-0.046, 0.073]	[-0.045, 0.073]	[-0.036, 0.060]	[-0.057, 0.081]	[-0.035, 0.054]
N	23967	10611	17625	22640	20274	29461
Bandwidth	4.01	1.41	2.51	3.62	3.05	6.53
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT
Panel D: Adjusted optimal bandwidths for each specification, CCT-procedure						
	(19)	(20)	(21)	(22)	(23)	(24)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.019	0.011	0.003	0.010	0.001	0.016
95% confidence interval (robust)	[-0.025, 0.062]	[-0.073, 0.096]	[-0.085, 0.091]	[-0.058, 0.079]	[-0.103, 0.105]	[-0.046, 0.078]
N	16738	6557	10373	14448	12645	22713
Bandwidth	2.37	0.83	1.38	1.98	1.70	3.64
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT

Notes: Table shows estimated incumbency advantage using local polynomial regressions within various bandwidths. Confidence intervals in panels A and B use standard errors clustered at municipality level.Panels C and D use the same main and bias bandwidths. Unit of observation is a candidate *i* at year *t*. Sample includes only small elections in which at most 2422 votes were given.

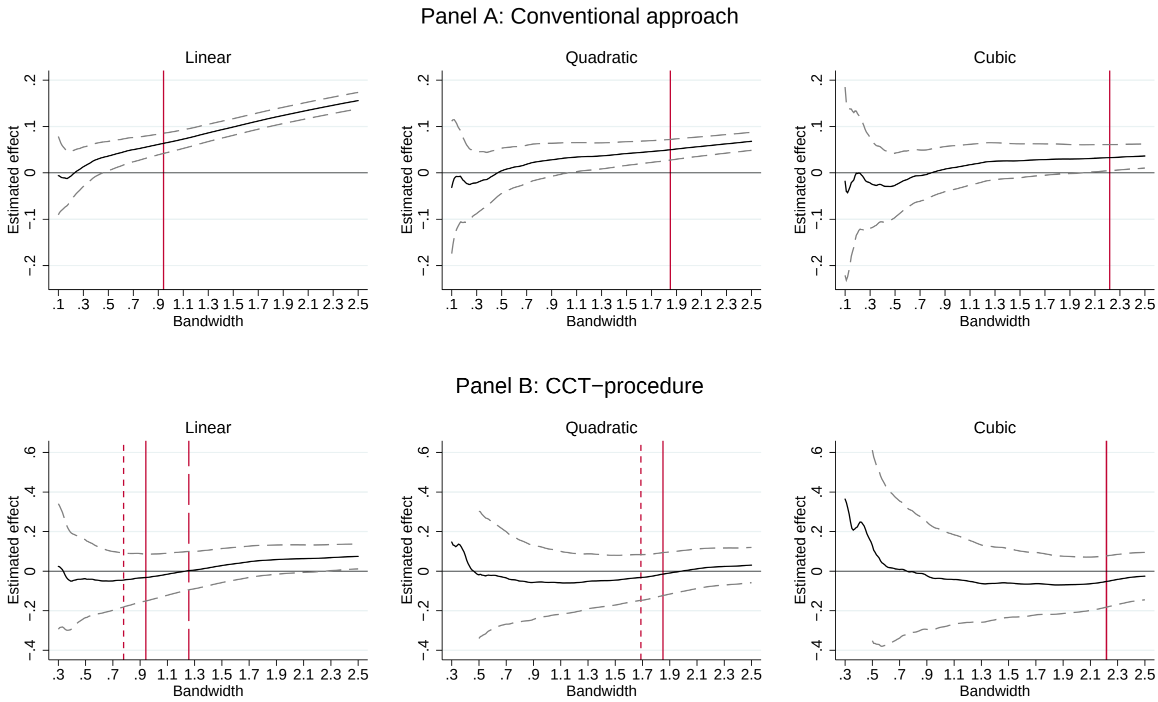
Table E4. RDD results for large municipalities.

Outcome: Elected next election						
Panel A: Bandwidth optimized for local linear specification						
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear		Quadratic		Cubic	
Elected	0.051	0.064	0.010	0.024	-0.026	-0.007
95% confidence interval (clustered)	[0.019, 0.082]	[0.036, 0.091]	[-0.038, 0.058]	[-0.020, 0.067]	[-0.090, 0.038]	[-0.063, 0.049]
N	17665	22917	17665	22917	17665	22917
Bandwidth	0.62	1.11	0.62	1.11	0.62	1.11
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	0.010	0.028	-0.035	-0.026	-0.031	-0.039
95% confidence interval (clustered)	[-0.035, 0.056]	[-0.012, 0.067]	[-0.103, 0.034]	[-0.086, 0.035]	[-0.129, 0.067]	[-0.121, 0.043]
N	8945	11344	8945	11344	8945	11344
Bandwidth	0.31	0.55	0.31	0.55	0.31	0.55
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT
Panel C: Bandwidths optimized for each specification, CCT-procedure						
	(13)	(14)	(15)	(16)	(17)	(18)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.010	0.024	0.026	0.037	0.016	0.041
95% confidence interval (robust)	[-0.035, 0.055]	[-0.016, 0.063]	[-0.013, 0.065]	[0.005, 0.068]	[-0.030, 0.061]	[0.012, 0.070]
N	17665	22917	42757	64160	50079	88588
Bandwidth	0.62	1.11	1.38	2.12	1.60	4.00
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT
Panel D: Adjusted optimal bandwidths for each specification, CCT-procedure						
	(19)	(20)	(21)	(22)	(23)	(24)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	-0.029	-0.016	-0.017	0.014	-0.035	0.023
95% confidence interval (robust)	[-0.094, 0.036]	[-0.071, 0.040]	[-0.075, 0.041]	[-0.031, 0.058]	[-0.104, 0.034]	[-0.017, 0.062]
N	9939	12571	20183	32711	24196	63415
Bandwidth	0.35	0.44	0.71	1.09	0.84	2.09
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT

Notes : Table shows estimated incumbency advantage using local polynomial regressions within various bandwidths. Confidence intervals in panels A and B use standard errors clustered at municipality level.Panels C and D use the same main and bias bandwidths. Unit of observation is a candidate *i* at year *t* . Sample includes only large elections in which more than 2422 voters voted.

Robustness test #4: Heterogeneity in the personal incumbency effect

Figure E1: This figure shows RDD point estimates and their 95 % confidence intervals for a wide range of bandwidths, obtained using only those party-lists that were involved in the lotteries. When these party-lists are used, increasing the bandwidths adds new candidates from the same lists, but does not add new lists or municipalities to the sample. The reason for reporting these results is that, besides the bias caused by the potentially incorrect linear approximation, the point estimates may increase due to heterogeneity in the personal incumbency effect across municipalities (and thus party-lists). Our baseline RDD may identify the effect for a different set of municipalities than what we have in the experimental sample. Moreover, we are in practice pooling many different thresholds located for example at different absolute number of votes to be located at the same normalized zero location in the forcing variable. In this exercise we are pooling exactly the same thresholds in both the experimental and RD sample. In Figure E1, we report the results both using the conventional approach (Panel A) and the CCT-procedure (Panel B) with the bias bandwidth fixed to the RD effect bandwidth. The findings reported below do not support the explanation of heterogeneous treatment effects, as the patterns that we find here are similar to those reported in the main text of HMSTT (Figure 2).

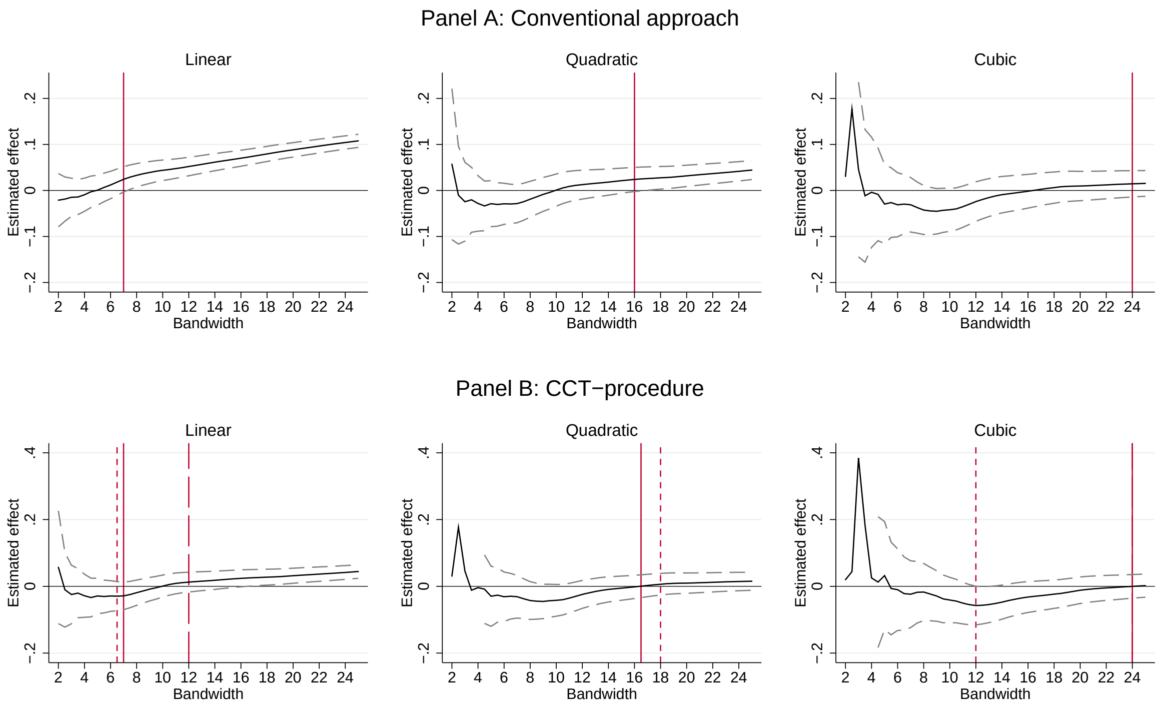


Notes: The graph displays the point estimates of incumbency advantage for various bandwidths using conventional approach (Panel A) and CCT-procedure (Panel B) with the same RD effect and bias bandwidth. Dashed lines mark the 95 % confidence intervals. In some of the figures, we do not display the confidence intervals for the smallest bandwidths in order to keep the scale of y-axes the same and thus the figures comparable. Red solid vertical line marks the optimal bandwidth chosen using IK implementation. Long-dashed vertical line marks the optimal CCT bandwidth and short-dashed line marks the adjusted CCT bandwidth. To keep the x-axes comparable, the (MSE-optimal and adjusted) CCT-bandwidths are shown only if they are smaller than 2.5. The sample includes only candidates from party lists that have lotteries.

Figure E1. RDD estimates using only party lists with lotteries.

Robustness test #5: Alternative definitions for the forcing variable.

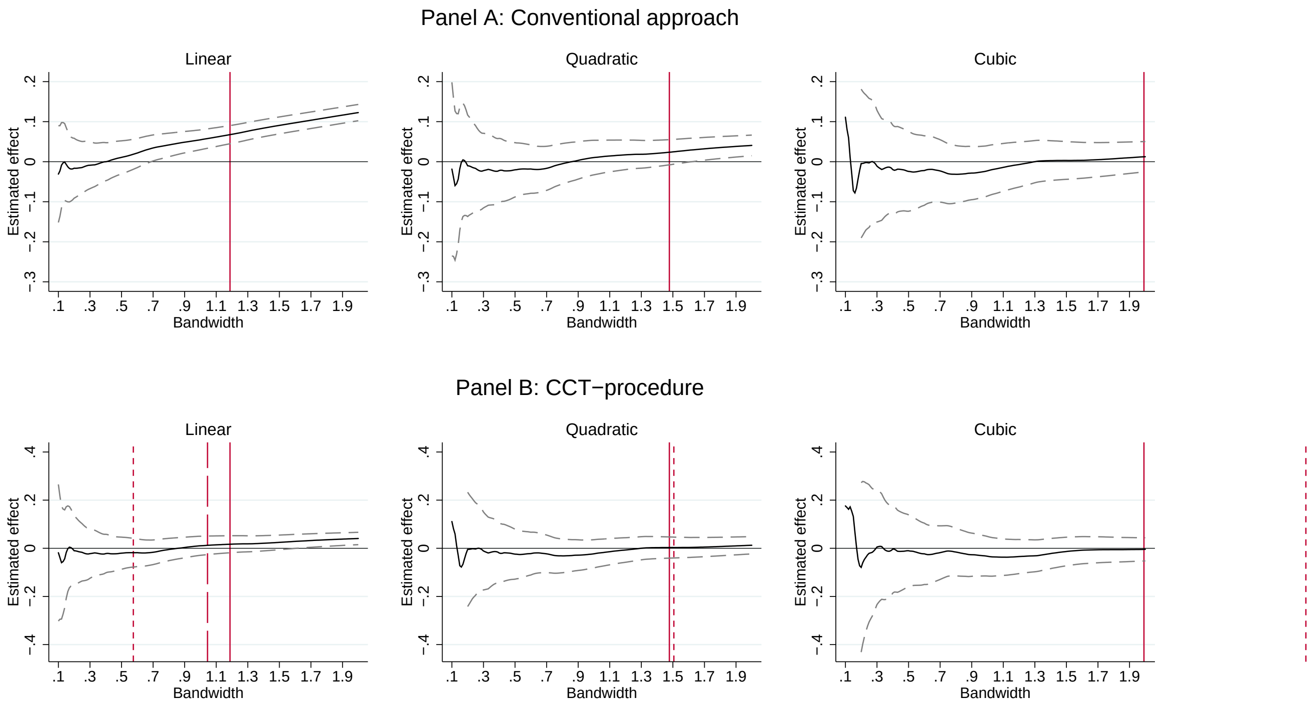
Figure E2: This figure reports RDD results when a non-scaled version of our forcing variable is used. The forcing variable is defined as in the main text of HMSTT, but is not scaled with the total number of votes the party got. We display the RDD estimates for linear, quadratic and cubic local polynomial specifications, separately for the conventional approach and the CCT-procedure. As the figure shows, the results that we obtain using this alternative forcing variable echo our baseline RDD results. The local linear polynomial produces biased results, but the higher order polynomials and bandwidths smaller than optimal work better. As Panel B shows, the bias-correction procedure of Calonico et al. (2014a) works well, especially if the MSE-optimal bandwidths are adjusted with the shrinkage factor suggested by Calonico et al. (2016a).



Notes: The graph displays the point estimates of incumbency advantage for various bandwidths using conventional approach (Panel A) and CCT-procedure (Panel B) with the same RD effect and bias bandwidth. Dashed lines mark the 95 % confidence intervals. In some of the figures, we do not display the confidence intervals for the smallest bandwidths in order to keep the scale of y-axes the same and thus the figures comparable. Red solid vertical line marks the optimal bandwidth chosen using IK implementation. Long-dashed vertical line marks the optimal CCT bandwidth and short-dashed line marks the adjusted CCT bandwidth. To keep the x-axes comparable, the (MSE-optimal and adjusted) CCT-bandwidths are shown only if they are smaller than 24. The forcing variable is as in the main text but not scaled with the total number of votes the party got.

Figure E2. RDD estimates using absolute vote margin, measured in number of votes, as the forcing variable.

Figure E3: This figure reports RDD results when another alternative version of our forcing variable is used. For this figure we define the cutoff as the number of votes of the first non-elected (last elected) candidate of the ordered party list for the elected (non-elected) candidates. The forcing variable is then the distance from this cutoff multiplied by 100 and divided by the number of party's votes. As the figure shows, the results echo our baseline RDD results. Moreover, as Panel B shows, the bias-correction procedure of Calonico et al. (2014a) works well, especially if the MSE-optimal bandwidths are adjusted with the shrinkage factor suggested by Calonico et al. (2016a).

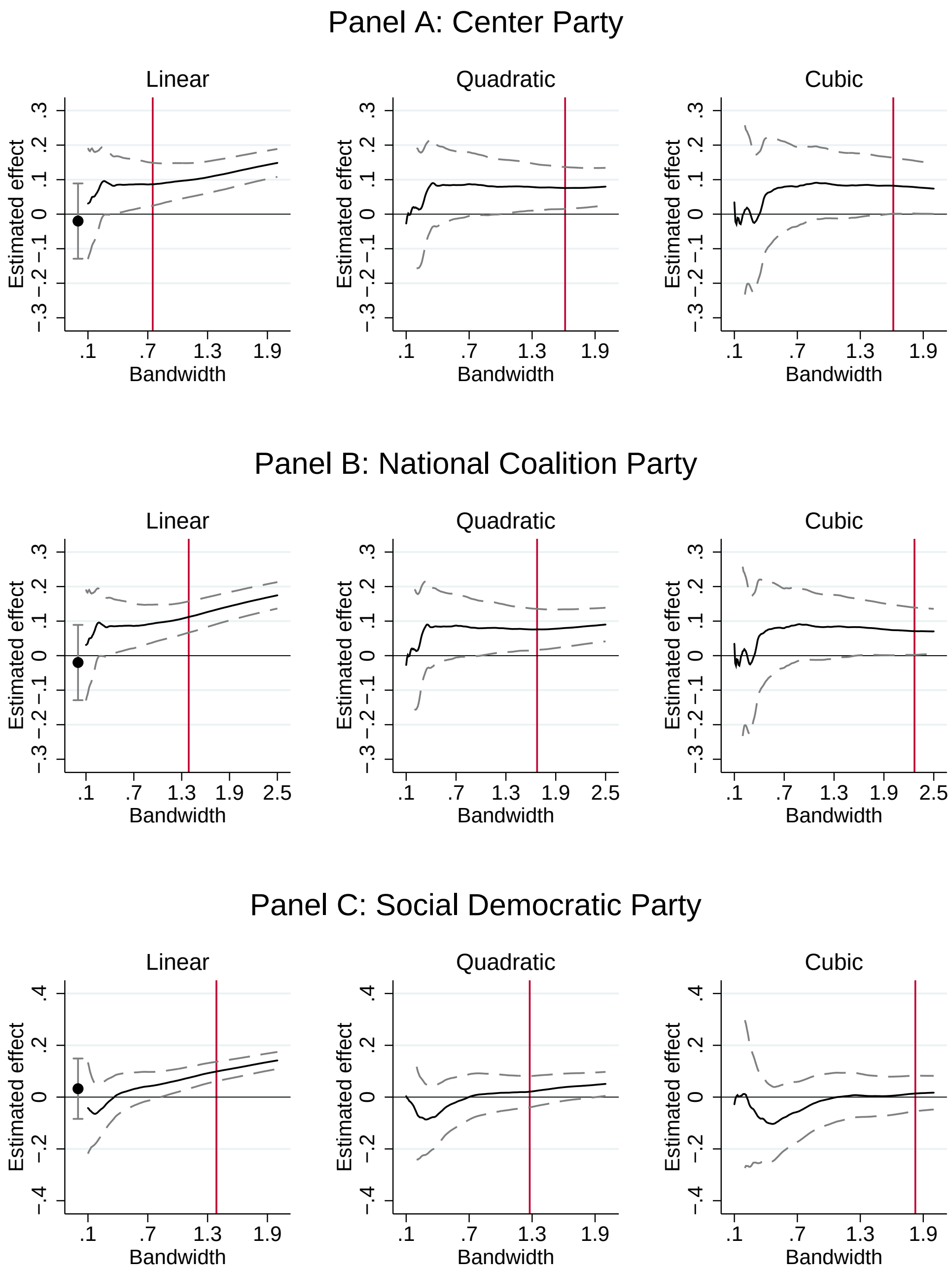


Notes: The graph displays the point estimates of incumbency advantage for various bandwidths using conventional approach (Panel A) and CCT-procedure (Panel B) with the same RD effect and bias bandwidth. Dashed lines mark the 95 % confidence intervals. In some of the figures, we do not display the confidence intervals for the smallest bandwidths in order to keep the scale of y-axes the same and thus the figures comparable. Red solid vertical line marks the optimal bandwidth chosen using IK implementation. Long-dashed vertical line marks the optimal CCT bandwidth and short-dashed line marks the adjusted CCT bandwidth. To keep the x-axes comparable, the (MSE-optimal and adjusted) CCT-bandwidths are shown only if they are smaller than 2. The forcing variable is then the distance from this cutoff multiplied by 100 and divided by the number of party's votes.

Figure E3. RDD estimates using the distance to the first non-elected (or last elected) candidate as the forcing variable.

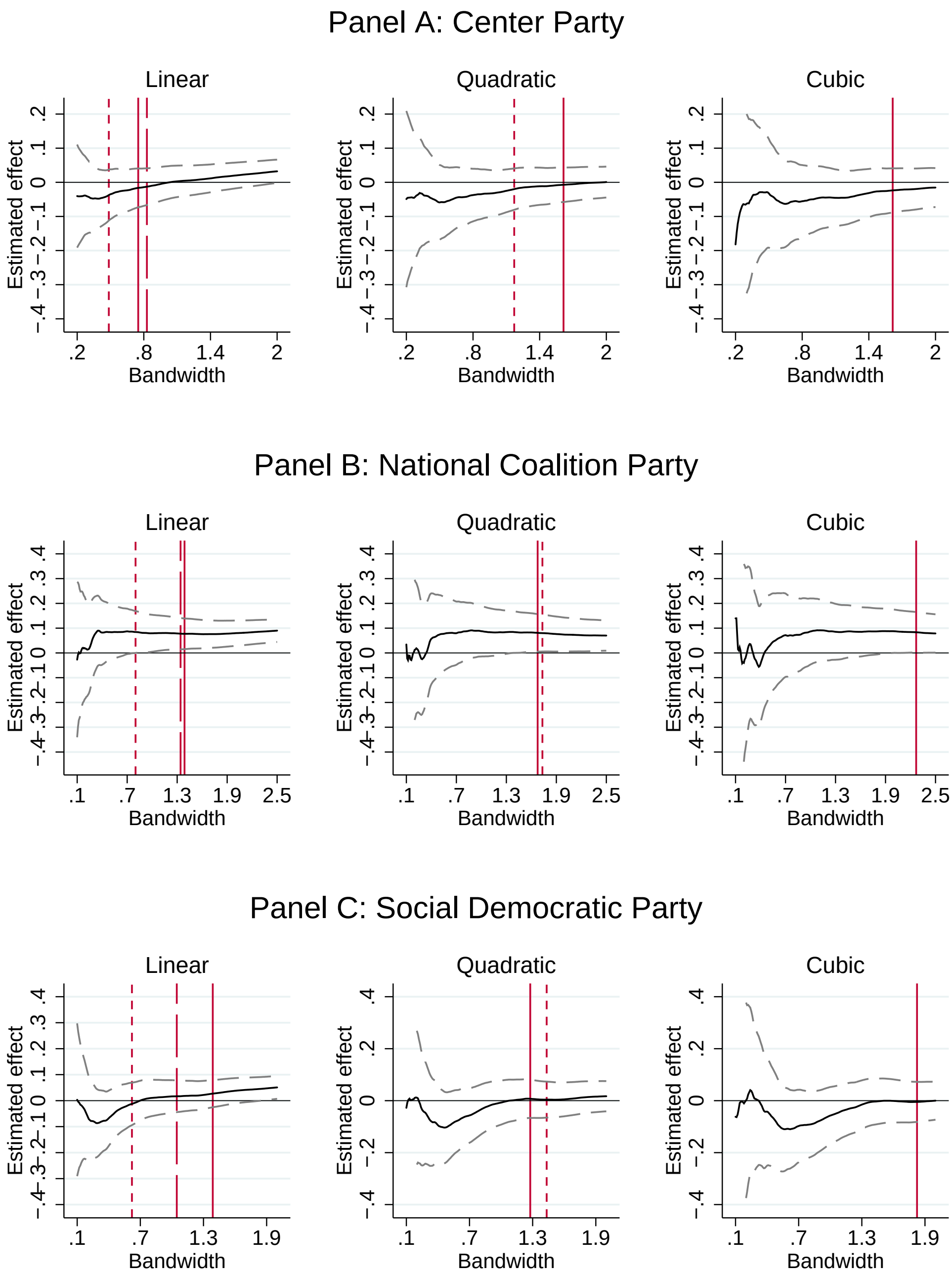
Robustness test #6: Heterogeneity in the effect between parties.

Figures E4 and E5: These figure reports graphically the RDD results separately for each of the three large parties (Panel A: Center Party, Panel B: National Coalition Party and Panel C: Social Democratic Party). Figure E4 shows results from conventional RDD estimations and Figure E5 reports the estimates obtained using CCT-procedure. The graphs allow us to study whether there is heterogeneity in the effect between the parties. Our motivation to look at such heterogeneity is that it could be an alternative explanation for the disparity between the experimental estimate and non-experimental RDD estimates. Suppose, for example, that there is no incumbency advantage within party A but a positive advantage within party B. Then if party A is more often involved in lotteries and if for some reason party B is overrepresented in the RDD samples (that are based on larger bandwidths), we might observe that the experimental estimate is zero and that RDD estimates produce a positive effect, especially when larger bandwidths are used. Figures E4 and E5 allow us to rule out such explanations. It seems that there is no substantial heterogeneity in the within party personal incumbency advantage between parties. As Figure E5 shows, the bias-correction procedure of Calonico et al. (2014a) works relatively well here, especially if the MSE-optimal bandwidths are adjusted with the shrinkage factor suggested by Calonico et al. (2016a)



Notes: The graph displays the point estimates of incumbency advantage for various bandwidths. Dashed lines show the 95 % confidence intervals. In some of the figures, we do not display the confidence intervals for the smallest bandwidths in order to keep the scale of the y-axes the same and thus the figures comparable. Red vertical line marks the optimal bandwidth chosen using IK implementation. The figure for linear specification also displays the estimate from the lottery sample and its 95 % confidence interval.

Figure E4. RDD estimates for different parties, conventional approach.



Notes: The graph displays the point estimates of incumbency advantage for various bandwidths. Dashed lines show the 95 % confidence intervals. In some of the figures, we do not display the confidence intervals for the smallest bandwidths in order to keep the scale of y-axes the same and thus the figures comparable. Red solid vertical line marks the optimal bandwidth chosen using IK implementation. Long-dashed vertical line marks the optimal CCT bandwidth and short-dashed line marks the adjusted CCT bandwidth. To keep the x-axes comparable within panels, the (MSE-optimal and adjusted) CCT-bandwidths are shown only if they are smaller than 2 (Panels A and C) or 2.5 (Panel B).

Figure E5. RDD estimates for different parties, CCT-procedure.

Robustness test #7: Excluding from the sample those who do not rerun

Tables E5 and E6: These tables report RDD results for a sample from which those who do not rerun are excluded. Table E5 reports the results for our main outcome, the effect of getting elected at period *t* on getting elected at period *t+1*. In Table E6, we look at an alternative outcome, incumbency advantage in vote share *t+1*. As we reported earlier (in Appendix B), the experimental estimates suggest no effect on these outcome variables when the sample from which those who do not rerun are excluded. Our motivation to report these results is that the previous literature is mixed on how those who do not rerun should be treated: For instance, Uppal (2010) report the results for a sample that includes all candidates and for a sample that only includes those who rerun, whereas de Magalhaes (2014) argues in favor of including all the candidates.

We again find that the standard implementation (local linear with IK optimal bandwidth) of RDD generates a positive and significant effect in both tables. We also find that undersmoothing appears to work (with one exception in Table E5, Panel B), and that the use of higher degree local polynomials without adjusting the bandwidth reproduces the experimental estimate in the sense that we do no reject the null hypothesis of no effect. These insignificant findings are largely, but not in each case, due to greater standard errors, as the estimated effects do not systematically become closer to zero as the more flexible approaches are used.

In Table E5, CCT-procedure suggests that there could be a small and statistically significant effect on getting elected at *t+1*. However, most of these estimates lose their statistical significance once we adjust the bandwidths following Calonico et al. (2016). The estimated effects are mostly smaller, but the conclusion of a zero effect is largely due to increased standard errors. Table E6 shows that, again, the local linear RDD with IK and CCT optimal bandwidths generates a positive and significant effect. However, both richer polynomials and the CCT-procedure recover the experimental estimate, irrespectively of whether the bandwidths are adjusted or not.

Table E5. RDD estimates using rerunners only, elected next election.

Outcome: Elected next election						
Panel A: Bandwidth optimized for local linear specification						
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear		Quadratic		Cubic	
Elected	0.067	0.075	0.051	0.053	0.037	0.043
95% confidence interval (clustered)	[0.026, 0.109]	[0.038, 0.112]	[-0.010, 0.111]	[-0.002, 0.108]	[-0.047, 0.121]	[-0.030, 0.115]
N	12058	15079	12058	15079	12058	15079
Bandwidth	0.54	0.69	0.54	0.69	0.54	0.69
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	0.048	0.057	0.034	0.035	0.056	0.034
95% confidence interval (clustered)	[-0.010, 0.107]	[0.006, 0.109]	[-0.055, 0.124]	[-0.044, 0.114]	[-0.077, 0.190]	[-0.077, 0.144]
N	6209	7745	6209	7745	6209	7745
Bandwidth	0.27	0.34	0.27	0.34	0.27	0.34
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT
Panel C: Bandwidths optimized for each specification, CCT-procedure						
	(13)	(14)	(15)	(16)	(17)	(18)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.051	0.053	0.059	0.056	0.060	0.051
95% confidence interval (robust)	[-0.009, 0.110]	[0.001, 0.105]	[0.013, 0.105]	[0.016, 0.097]	[0.012, 0.108]	[0.014, 0.087]
N	12058	15079	31503	39265	42257	56704
Bandwidth	0.54	0.69	1.47	1.90	2.10	3.62
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT
Panel D: Adjusted optimal bandwidths for each specification, CCT-procedure						
	(19)	(20)	(21)	(22)	(23)	(24)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.030	0.038	0.043	0.052	0.045	0.061
95% confidence interval (robust)	[-0.056, 0.116]	[-0.035, 0.112]	[-0.025, 0.111]	[-0.006, 0.110]	[-0.025, 0.115]	[0.010, 0.111]
N	7017	8783	16780	21631	24365	39851
Bandwidth	0.31	0.39	0.77	0.99	1.12	1.93
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT

Notes: Table shows estimated incumbency advantage using local polynomial regressions within various bandwidths. Sample includes only rerunning candidates. Confidence intervals in panels A and B use standard errors clustered at municipality level.Panels C and D use the same main and bias bandwidths. Unit of observation is a candidate *i* at year *t*.

Table E6. RDD estimates using rerunners only, vote share next election.

Outcome: Vote share next election						
Panel A: Bandwidth optimized for local linear specification						
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear		Quadratic		Cubic	
Elected	0.049	0.049	0.047	0.047	0.049	0.052
95% confidence interval (clustered)	[0.002, 0.096]	[0.002, 0.097]	[-0.017, 0.111]	[-0.019, 0.113]	[-0.037, 0.134]	[-0.037, 0.141]
N	16668	15697	16668	15697	16668	15697
Bandwidth	0.76	0.72	0.76	0.72	0.76	0.72
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	0.058	0.060	0.037	0.026	-0.028	-0.028
95% confidence interval (clustered)	[-0.003, 0.118]	[-0.002, 0.122]	[-0.053, 0.127]	[-0.066, 0.119]	[-0.145, 0.089]	[-0.148, 0.093]
N	16668	15697	16668	15697	16668	15697
Bandwidth	0.38	0.36	0.38	0.36	0.38	0.36
Bandwidth selection method	0.5 * IK	0.5 * CCT	0.5 * IK	0.5 * CCT	0.5 * IK	0.5 * CCT
Panel C: Bandwidths optimized for each specification, CCT-procedure						
	(13)	(14)	(15)	(16)	(17)	(18)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.047	0.047	0.028	0.029	0.038	0.030
95% confidence interval (robust)	[-0.028, 0.122]	[-0.030, 0.124]	[-0.041, 0.097]	[-0.037, 0.095]	[-0.031, 0.106]	[-0.033, 0.092]
N	16668	15697	35817	39168	49438	55966
Bandwidth	0.76	0.72	1.70	1.89	2.72	3.51
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT
Panel D: Adjusted optimal bandwidths for each specification, CCT-procedure						
	(19)	(20)	(21)	(22)	(23)	(24)
	Linear		Quadratic		Cubic	
Elected (bias-corrected)	0.051	0.046	0.052	0.048	0.056	0.038
95% confidence interval (robust)	[-0.044, 0.146]	[-0.051, 0.143]	[-0.038, 0.143]	[-0.039, 0.134]	[-0.033, 0.145]	[-0.042, 0.117]
N	9709	9129	19329	21566	31212	38886
Bandwidth	0.43	0.41	0.89	0.99	1.45	1.87
Bandwidth selection method	IK	CCT	IK	CCT	IK	CCT

Notes : Table shows estimated incumbency advantage using local polynomial regressions within various bandwidths. Sample includes only rerunning candidates. Confidence intervals in panels A and B use standard errors clustered at municipality level.Panels C and D use the same main and bias bandwidths. Unit of observation is a candidate *i* at year *t* .

Robustness test #8: New rdrobust package

We have re-estimated the most relevant specifications of our analysis using the new MSE- and CER-bandwidths, made available by the updated version of rdrobust software (see Calonico et al. 2016b). The CER-optimal bandwidth is based on a higher-order Edgeworth expansion. This bandwidth optimizes coverage error but does not necessarily have desirable properties for point estimation. The updated software also allows for clustering when calculating the standard errors and the bandwidths.

Tables E7 and E8: Table E7 reports conventional point estimates in Panel A, bias-corrected point estimates in Panels B and C, and confidence intervals allowing for clustering at the municipality level. In Panel A of Table E7, we use the conventional approach and the bandwidth is selected optimally either for the local linear specification (columns (1)-(4)) or the local quadratic specification (columns (5) and (6)) using the new MSE- and CER-bandwidths. Panels B and C report results obtained using the CCT-procedure. In Panel B, we estimate the bandwidths for the RDD effect and bias separately, while these two are fixed to be equal in Panel C. The results largely echo our earlier findings and support our earlier conclusions. In particular, fitting local polynomials within optimal bandwidths may lead to misleading results if the bandwidths are too wide. The new implementation of the MSE-optimal bandwidth is similar to the CCT implementation in the older version of rdrobust software. The results that the new MSE implementation produces are therefore similar to what we report for the CCT implementation. More generally, it seems that the exact way of implementing the MSE-optimal bandwidth is less relevant than following the recommendations of Calonico et al. (2016a); what reproduces the experimental estimate in our data is fitting polynomials of degree p+1 within the optimal bandwidth for p or setting the RDD effect and bias bandwidths equal (Panel C). We also allowed for different bandwidths for the treatment and the control groups; that did not substantially affect the results (not reported). Table E8 replicates Table E7 but reports non-clustered (but heteroscedastic-robust) standard errors. As can be seen, the results are similar, if no clustering is used.

Table E7. RDD estimates with new MSE and CER-optimal bandwidths (clustered standard errors).

Panel A: Conventional approach						
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear		Quadratic (bandwidth for p = 1)		Quadratic (bandwidth for p = 2)	
Elected	0.051	0.038	0.021	0.006	0.059	0.041
95% confidence interval (clustered)	[0.026, 0.077]	[0.006, 0.069]	[-0.019, 0.061]	[-0.041, 0.054]	[0.039, 0.080]	[0.016, 0.067]
N	26463	18804	26463	18804	80971	57225
R^2	0.05	0.03	0.05	0.03	0.17	0.12
Bandwidth	0.73	0.52	0.73	0.52	2.18	1.48
Bandwidth implementation	MSE	CER	MSE	CER	MSE	CER
Panel B: CCT-procedure with optimal bandwidths						
	(7)	(8)	(9)	(10)	(11)	(12)
	Linear		Quadratic (bandwidth for p = 1)		Quadratic (bandwidth for p = 2)	
Elected (bias-corrected)	0.045	0.034	-0.005	-0.024	0.055	0.040
95% confidence interval (clustered)	[0.019, 0.071]	[0.003, 0.065]	[-0.058, 0.048]	[-0.091, 0.042]	[0.034, 0.076]	[0.014, 0.067]
N	14506	10415	14506	10415	41983	27580
RD effect bandwidth	0.73	0.52	0.73	0.52	2.18	1.48
Bias bandwidth	3.01	3.01	3.01	3.01	6.34	6.34
Bandwidth implementation	MSE	CER	MSE	CER	MSE	CER
Panel C: CCT-procedure with RD effect bandwidth equal to bias bandwidth						
	(13)	(14)	(15)	(16)	(17)	(18)
	Linear		Quadratic (bandwidth for p = 1)		Quadratic (bandwidth for p = 2)	
Elected (bias-corrected)	0.021	0.006	-0.005	-0.024	0.033	0.026
95% confidence interval (clustered)	[-0.018, 0.060]	[-0.040, 0.053]	[-0.058, 0.048]	[-0.091, 0.042]	[0.004, 0.061]	[-0.010, 0.062]
N	14506	10415	14506	10415	41983	27580
RD effect bandwidth	0.73	0.52	0.73	0.52	2.18	1.48
Bias bandwidth	0.73	0.52	0.73	0.52	2.18	1.48
Bandwidth implementation	MSE	CER	MSE	CER	MSE	CER

Notes: Table shows estimated incumbency advantage using local polynomial regressions within various bandwidths. All estimations use a triangular kernel. Confidence intervals account for clustering at municipality level. Unit of observation is a candidate *i* at year *t*. The MSE bandwidth is a newer implementation of the estimation of the MSE-optimal bandwidth choice (see Calonico et al. 2016b).

Table E8. RDD estimates with new MSE and CER-optimal bandwidths (non-clustered standard errors).

Outcome: Elected next election						
Panel A: Conventional approach						
	(1)	(2)	(3)	(4)	(5)	(6)
	Linear		Quadratic (bandwidth for p = 1)		Quadratic (bandwidth for p = 2)	
Elected	0.051	0.028	0.020	-0.012	0.060	0.034
95% confidence interval (non-clustered)	[0.026, 0.076]	[-0.006, 0.062]	[-0.017, 0.058]	[-0.067, 0.043]	[0.039, 0.081]	[0.005, 0.064]
N	26221	14404	26221	14404	81696	42090
R^2	0.05	0.02	0.05	0.02	0.17	0.09
Bandwidth	0.72	0.40	0.72	0.40	2.20	1.11
Bandwidth selection method	MSE	CER	MSE	CER	MSE	CER
Panel B: Bias-correction with optimal bandwidths						
	(7)	(8)	(9)	(10)	(11)	(12)
	Linear		Quadratic (bandwidth for p = 1)		Quadratic (bandwidth for p = 2)	
Elected (bias-corrected)	0.045	0.026	-0.005	-0.026	0.056	0.034
95% confidence interval (non-clustered)	[0.020, 0.070]	[-0.008, 0.060]	[-0.059, 0.048]	[-0.109, 0.056]	[0.035, 0.077]	[0.005, 0.063]
N	26221	14404	26221	14404	81696	42090
Bandwidth	0.72	0.40	0.72	0.40	2.20	1.11
Bias bandwidth	3.05	3.05	3.05	3.05	6.53	6.53
Bandwidth selection method	MSE	CER	MSE	CER	MSE	CER
Panel C: Bias-correction with main bandwidth equal to pilot bandwidth						
	(13)	(14)	(15)	(16)	(17)	(18)
	Linear		Quadratic (bandwidth for p = 1)		Quadratic (bandwidth for p = 2)	
Elected (bias-corrected)	0.020	-0.012	-0.005	-0.026	0.033	0.018
95% confidence interval (non-clustered)	[-0.017, 0.058]	[-0.067, 0.044]	[-0.059, 0.048]	[-0.109, 0.056]	[0.006, 0.060]	[-0.022, 0.058]
N	26221	14404	26221	14404	81696	42090
Bandwidth	0.72	0.40	0.72	0.40	2.20	1.11
Bias bandwidth	0.72	0.40	0.72	0.40	2.20	1.11
Bandwidth selection method	MSE	CER	MSE	CER	MSE	CER

Notes: Table shows estimated incumbency advantage using local polynomial regressions within various bandwidths. All estimations use a triangular kernel. Confidence intervals are computed using heteroskedasticity-robust standard errors. Unit of observation is a candidate *i* at year *t*.

Appendix F: Supplementary information to Section 4.3 (When is RDD as good as randomly assigned?)

This appendix reports the means tests of covariate balance within small bandwidths near the cutoff as well as a brief analysis of when RDD is as good as randomly assigned using the approach proposed by Cattaneo et al. (2015).

Means tests of covariate balance within small bandwidths near the cutoff

The tests reported below do not control for the slopes (or curvature) of the forcing variable nearby the cutoff. They are not tests of whether the covariates develop smoothly over the cutoff, but rather tests for whether the treatment is as good as randomly assigned. The 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.

Table F1 and **F2:** Table F1 looks at the covariate balance of candidate characteristics. It reports the means of the candidate characteristics for small bandwidths on both sides of the cutoff as well as a *t*-test for the difference of the means. For example, when incumbency status (elected at *t*-1) is used, we find that bandwidths 0.04 or smaller are as-good-as-random at the 5% significance level (923 observations). Based on a minimum *p*-value criterion among all the covariates (but not correcting for multiple testing), it seems that bandwidths 0.02 or smaller would be as-good-as random at the 5% significance level (128 observations). These numbers are obtained by starting from the zero bandwidth and widening the bandwidth until the first statistically significant coefficient is found. This is a conservative approach in the sense that if we started from wider bandwidths and decreased their length until no significant differences are found, we would get somewhat larger bandwidth estimates. For example, based on Table F1, a bandwidth of 0.05 would be as-good-as-random (but 0.10 or larger would not). Table F2 reproduces the analysis of Table F1 for municipality-level covariates. As the table shows, they are balanced, as they should be by construction.

Table F1. Covariate balance within small bandwidths (candidate characteristics).

Bandwidth = 0.01

Variable	Elected (N = 37) N	Mean	Std. dev.	Not elected (N = 38) N	Mean	Std. dev.	Difference
Vote share	37	0.49	0.20	38	0.46	0.18	0.02
Number of votes	37	291	230	38	282	208	9
Female	37	0.49	0.51	38	0.47	0.51	0.01
Age	37	47.62	9.91	38	49.74	11.86	-2.12
Incumbent	37	0.41	0.50	38	0.55	0.50	-0.15
Municipal employee	37	0.43	0.50	38	0.37	0.49	0.06
Wage income	27	34676	17829	28	35972	21286	-1296
Capital income	27	1670	6528	28	1426	3137	244
High professional	37	0.49	0.51	38	0.34	0.48	0.14
Entrepreneur	37	0.08	0.28	38	0.05	0.23	0.03
Student	37	0.00	0.00	38	0.03	0.16	-0.03
Unemployed	37	0.03	0.16	38	0.00	0.00	0.03
University degree	33	0.30	0.47	32	0.22	0.42	0.08

Bandwidth = 0.05

Variable	Elected (N = 729) N	Mean	Std. dev.	Not elected (N = 761) N	Mean	Std. dev.	Difference
Vote share	729	0.90	0.44	761	0.85	0.43	0.05
Number of votes	729	102	111	761	102	111	0
Female	729	0.41	0.49	761	0.37	0.48	0.04
Age	729	47.04	11.79	761	47.01	12.43	0.03
Incumbent	729	0.39	0.49	761	0.34	0.48	0.05
Municipal employee	729	0.30	0.46	761	0.26	0.44	0.03
Wage income	542	27672	17445	576	27397	18042	275
Capital income	542	5547	65197	576	1884	5972	3663
High professional	729	0.26	0.44	761	0.27	0.44	-0.01
Entrepreneur	729	0.17	0.37	761	0.18	0.38	-0.01
Student	729	0.02	0.15	761	0.03	0.17	-0.01
Unemployed	729	0.05	0.21	761	0.03	0.17	0.02
University degree	597	0.19	0.39	641	0.17	0.38	0.02

Bandwidth = 0.10

Variable	Elected (N = 1778) N	Mean	Std. dev.	Not elected (N = 1906) N	Mean	Std. dev.	Difference
Vote share	1778	1.02	0.53	1906	0.95	0.50	0.07
Number of votes	1778	90	104	1906	87	101	3
Female	1778	0.38	0.49	1906	0.39	0.49	0.00
Age	1778	46.45	11.79	1906	46.47	11.94	-0.03
Incumbent	1778	0.36	0.48	1906	0.31	0.46	0.05**
Municipal employee	1778	0.27	0.45	1906	0.27	0.44	0.01
Wage income	1334	26729	17385	1437	27470	20890	-741
Capital income	1334	3635	42060	1437	3080	24308	556
High professional	1778	0.25	0.43	1906	0.25	0.43	0.00
Entrepreneur	1778	0.19	0.39	1906	0.19	0.39	0.00
Student	1778	0.02	0.14	1906	0.03	0.18	-0.01**
Unemployed	1778	0.04	0.20	1906	0.03	0.18	0.01
University degree	1466	0.19	0.39	1576	0.17	0.38	0.02

Bandwidth = 0.20

Variable	Elected (N = 3508) N	Mean	Std. dev.	Not elected (N = 3891) N	Mean	Std. dev.	Difference
Vote share	3508	1.11	0.58	3891	0.98	0.54	0.12
Number of votes	3508	85	101	3891	81	96	5
Female	3508	0.36	0.48	3891	0.40	0.49	-0.03**
Age	3508	46.51	11.59	3891	46.17	12.02	0.34
Incumbent	3508	0.36	0.48	3891	0.29	0.45	0.07**
Municipal employee	3508	0.26	0.44	3891	0.27	0.44	0.00
Wage income	2610	26852	18849	2892	26414	21339	439
Capital income	2610	3068	30512	2892	3416	27048	-347
High professional	3508	0.24	0.43	3889	0.24	0.43	0.00
Entrepreneur	3508	0.21	0.40	3889	0.19	0.39	0.02
Student	3508	0.02	0.15	3889	0.03	0.18	-0.01*
Unemployed	3508	0.04	0.20	3889	0.04	0.19	0.00
University degree	2858	0.18	0.39	3201	0.17	0.38	0.01

Bandwidth = 0.30

Variable	Elected (N = 5172) N	Mean	Std. dev.	Not elected (N = 5879) N	Mean	Std. dev.	Difference
Vote share	5172	1.16	0.62	5879	0.98	0.56	0.18*
Number of votes	5172	86	109	5879	77	91	9
Female	5172	0.36	0.48	5879	0.39	0.49	-0.03**
Age	5172	46.55	11.66	5879	46.28	12.05	0.27
Incumbent	5172	0.38	0.49	5879	0.28	0.45	0.10*
Municipal employee	5172	0.26	0.44	5879	0.26	0.44	0.00
Wage income	3829	26867	19027	4355	26305	20454	562
Capital income	3829	3131	26071	4355	3256	23362	-125
High professional	5172	0.23	0.42	5877	0.24	0.43	-0.01
Entrepreneur	5172	0.21	0.41	5877	0.19	0.39	0.03
Student	5172	0.02	0.15	5877	0.03	0.17	-0.01
Unemployed	5172	0.04	0.20	5877	0.04	0.20	0.00
University degree	4232	0.18	0.38	4832	0.17	0.38	0.01

Bandwidth = 0.40

Variable	Elected (N = 6628) N	Mean	Std. dev.	Not elected (N = 7949) N	Mean	Std. dev.	Difference
Vote share	6628	1.19	0.63	7949	0.95	0.56	0.24**
Number of votes	6628	87	114	7949	76	89	11
Female	6628	0.36	0.48	7949	0.39	0.49	-0.03**
Age	6628	46.74	11.62	7949	46.19	12.18	0.55
Incumbent	6628	0.39	0.49	7949	0.26	0.44	0.13**
Municipal employee	6628	0.26	0.44	7949	0.25	0.43	0.01
Wage income	4909	26685	19495	5911	26144	19643	541
Capital income	4909	3187	23615	5911	2994	20907	193
High professional	6628	0.23	0.42	7947	0.24	0.42	-0.01
Entrepreneur	6628	0.21	0.41	7947	0.18	0.38	0.03*
Student	6628	0.02	0.15	7947	0.03	0.17	-0.01*
Unemployed	6628	0.04	0.20	7947	0.04	0.20	0.00
University degree	5402	0.18	0.38	6534	0.17	0.38	0.00

Bandwidth = 0.55

Variable	Elected (N = 8710) N	Mean	Std. dev.	Not elected (N = 11348) N	Mean	Std. dev.	Difference
Vote share	8710	1.23	0.65	11348	0.92	0.56	0.32**
Number of votes	8710	87	113	11348	74	85	13
Female	8710	0.37	0.48	11348	0.40	0.49	-0.03**
Age	8710	46.77	11.56	11348	46.06	12.22	0.71*
Incumbent	8710	0.40	0.49	11348	0.24	0.43	0.16**
Municipal employee	8710	0.26	0.44	11348	0.25	0.43	0.01
Wage income	6468	26480	19009	8416	26115	19255	365
Capital income	6468	3315	23909	8416	2782	18532	533
High professional	8710	0.23	0.42	11345	0.24	0.42	-0.01
Entrepreneur	8710	0.22	0.41	11345	0.17	0.38	0.05**
Student	8710	0.02	0.15	11345	0.03	0.18	-0.01**
Unemployed	8710	0.04	0.20	11345	0.04	0.21	0.00
University degree	7091	0.17	0.38	9323	0.18	0.38	0.00

Bandwidth = 0.70

Variable	Elected (N = 10632) N	Mean	Std. dev.	Not elected (N = 14888) N	Mean	Std. dev.	Difference
Vote share	10632	1.27	0.66	14888	0.87	0.56	0.40**
Number of votes	10632	87	112	14888	72	80	15
Female	10632	0.36	0.48	14888	0.40	0.49	-0.04**
Age	10632	46.84	11.50	14888	45.89	12.27	0.95**
Incumbent	10632	0.42	0.49	14888	0.21	0.41	0.21**
Municipal employee	10632	0.26	0.44	14888	0.25	0.43	0.01
Wage income	7876	26362	18766	10981	26203	26303	159
Capital income	7876	3505	23583	10981	2757	18799	748**
High professional	10631	0.22	0.42	14884	0.24	0.43	-0.01
Entrepreneur	10631	0.22	0.42	14884	0.17	0.38	0.06**
Student	10631	0.02	0.15	14884	0.03	0.18	-0.01**
Unemployed	10631	0.04	0.19	14884	0.04	0.21	-0.01
University degree	8634	0.17	0.37	12252	0.18	0.38	-0.01

Bandwidth = 1.00

Variable	Elected (N = 14138) N	Mean	Std. dev.	Not elected (N = 23320) N	Mean	Std. dev.	Difference
Vote share	14138	1.34	0.69	23320	0.76	0.56	0.58**
Number of votes	14138	88	114	23320	66	71	22*
Female	14138	0.36	0.48	23320	0.41	0.49	-0.05**
Age	14138	47.01	11.40	23320	45.56	12.29	1.44**
Incumbent	14138	0.44	0.50	23320	0.17	0.37	0.27**
Municipal employee	14138	0.26	0.44	23320	0.25	0.43	0.01
Wage income	10438	26413	18889	17166	25903	23685	509
Capital income	10438	3691	24916	17166	2447	16276	1244***
High professional	14136	0.22	0.42	23314	0.24	0.42	-0.01
Entrepreneur	14136	0.23	0.42	23314	0.16	0.37	0.07**
Student	14136	0.23	0.15	23314	0.04	0.19	-0.01**
Unemployed	14136	0.23	0.19	23314	0.05	0.21	-0.01**
University degree	11490	0.23	0.38	19252	0.18	0.38	-0.01

Notes : * and ** denote 5 % and 1 % statistical significance of difference in means respectively. The significance of differences is tested using t test adjusted for clustering by municipality. 0.55 bandwidth equals roughly the optimal bandwidth chosen using Imbens and Kalyanaraman's (2008) algorithm. Sample includes only candidates running in 1996-2008 elections. Lotteries have been excluded. In 1996 elections income data are available only for candidates who run also in 2000, 2004 and 2008 elections. Income is expressed in euros.

Table F1 (continued). Covariate balance within small bandwidths (candidate characteristics).

Bandwidth = 0.01	Elected (N = 37)			Not elected (N = 38)			
Variable	N	Mean	Std. dev.	N	Mean	Std. dev.	Difference
Coalition Party	37	0.32	0.47	38	0.34	0.48	-0.02
Social Democrats	37	0.32	0.47	38	0.32	0.47	0.01
Center Party	37	0.11	0.31	38	0.11	0.31	0.00
True Finns	37	0.00	0.00	38	0.00	0.00	0.00
Green Party	37	0.11	0.31	38	0.11	0.31	0.00
Socialist Party	37	0.08	0.28	38	0.08	0.27	0.00
Swedish Party	37	0.05	0.23	38	0.05	0.23	0.00
Christian Party	37	0.00	0.00	38	0.00	0.00	0.00
Other parties	37	0.00	0.00	38	0.00	0.00	0.00

Bandwidth = 0.05	Elected (N = 729)			Not elected (N = 761)			
Variable	N	Mean	Std. dev.	N	Mean	Std. dev.	Difference
Coalition Party	729	0.18	0.38	761	0.18	0.38	0.00
Social Democrats	729	0.27	0.44	761	0.28	0.45	-0.01
Center Party	729	0.40	0.49	761	0.40	0.49	0.01
True Finns	729	0.00	0.04	761	0.00	0.04	0.00
Green Party	729	0.02	0.14	761	0.02	0.14	0.00
Socialist Party	729	0.04	0.20	761	0.04	0.19	0.00
Swedish Party	729	0.07	0.26	761	0.07	0.25	0.00
Christian Party	729	0.00	0.05	761	0.00	0.06	0.00
Other parties	729	0.02	0.12	761	0.01	0.12	0.00

Bandwidth = 0.10	Elected (N = 1778)			Not elected (N = 1906)			
Variable	N	Mean	Std. dev.	N	Mean	Std. dev.	Difference
Coalition Party	1778	0.19	0.39	1906	0.18	0.39	0.00
Social Democrats	1778	0.24	0.43	1906	0.25	0.43	0.00
Center Party	1778	0.41	0.49	1906	0.41	0.49	0.00
True Finns	1778	0.00	0.06	1906	0.00	0.06	0.00
Green Party	1778	0.02	0.15	1906	0.02	0.15	0.00
Socialist Party	1778	0.06	0.23	1906	0.05	0.22	0.01
Swedish Party	1778	0.06	0.23	1906	0.05	0.23	0.00
Christian Party	1778	0.00	0.06	1906	0.00	0.07	0.00
Other parties	1778	0.02	0.14	1906	0.02	0.14	0.00

Bandwidth = 0.20	Elected (N = 3508)			Not elected (N = 3891)			
Variable	N	Mean	Std. dev.	N	Mean	Std. dev.	Difference
Coalition Party	3508	0.18	0.39	3891	0.19	0.39	-0.01
Social Democrats	3508	0.24	0.43	3891	0.25	0.43	-0.01
Center Party	3508	0.41	0.49	3891	0.39	0.49	0.01
True Finns	3508	0.01	0.07	3891	0.01	0.07	0.00
Green Party	3508	0.02	0.15	3891	0.02	0.15	0.00
Socialist Party	3508	0.06	0.23	3891	0.05	0.23	0.00
Swedish Party	3508	0.06	0.24	3891	0.06	0.24	0.00
Christian Party	3508	0.01	0.08	3891	0.01	0.08	0.00
Other parties	3508	0.02	0.14	3891	0.02	0.14	0.00

Bandwidth = 0.30	Elected (N = 5172)			Not elected (N = 5879)			
Variable	N	Mean	Std. dev.	N	Mean	Std. dev.	Difference
Coalition Party	5172	0.19	0.39	5879	0.19	0.40	-0.01
Social Democrats	5172	0.24	0.42	5879	0.25	0.43	-0.01
Center Party	5172	0.40	0.49	5879	0.38	0.49	0.02
True Finns	5172	0.01	0.08	5879	0.01	0.08	0.00
Green Party	5172	0.02	0.15	5879	0.02	0.15	0.00
Socialist Party	5172	0.06	0.23	5879	0.06	0.23	0.00
Swedish Party	5172	0.06	0.24	5879	0.06	0.23	0.00
Christian Party	5172	0.01	0.09	5879	0.01	0.10	0.00
Other parties	5172	0.02	0.14	5879	0.02	0.15	0.00

Bandwidth = 0.40	Elected (N = 6628)			Not elected (N = 7949)			
Variable	N	Mean	Std. dev.	N	Mean	Std. dev.	Difference
Coalition Party	6628	0.18	0.39	7949	0.19	0.39	-0.01
Social Democrats	6628	0.23	0.42	7949	0.26	0.44	-0.02
Center Party	6628	0.41	0.49	7949	0.38	0.48	0.03
True Finns	6628	0.01	0.09	7949	0.01	0.08	0.00
Green Party	6628	0.02	0.14	7949	0.03	0.16	0.00
Socialist Party	6628	0.06	0.24	7949	0.06	0.23	0.00
Swedish Party	6628	0.06	0.23	7949	0.05	0.23	0.00
Christian Party	6628	0.01	0.10	7949	0.01	0.10	0.00
Other parties	6628	0.02	0.15	7949	0.02	0.15	0.00

Bandwidth = 0.55	Elected (N = 8710)			Not elected (N = 11348)			
Variable	N	Mean	Std. dev.	N	Mean	Std. dev.	Difference
Coalition Party	8710	0.18	0.39	11348	0.20	0.40	-0.01
Social Democrats	8710	0.23	0.42	11348	0.26	0.44	-0.03
Center Party	8710	0.40	0.49	11348	0.36	0.48	0.04
True Finns	8710	0.01	0.09	11348	0.01	0.09	0.00
Green Party	8710	0.02	0.14	11348	0.03	0.16	-0.01
Socialist Party	8710	0.06	0.24	11348	0.06	0.23	0.00
Swedish Party	8710	0.06	0.24	11348	0.05	0.22	0.01
Christian Party	8710	0.01	0.10	11348	0.01	0.10	0.00
Other parties	8710	0.02	0.15	11348	0.02	0.15	0.00

Bandwidth = 0.70	Elected (N = 10632)			Not elected (N = 14888)			
Variable	N	Mean	Std. dev.	N	Mean	Std. dev.	Difference
Coalition Party	10632	0.18	0.38	14888	0.20	0.40	-0.02
Social Democrats	10632	0.23	0.42	14888	0.27	0.44	0.04
Center Party	10632	0.41	0.49	14888	0.35	0.48	0.06
True Finns	10632	0.01	0.09	14888	0.01	0.08	0.00
Green Party	10632	0.02	0.14	14888	0.03	0.17	-0.01
Socialist Party	10632	0.06	0.24	14888	0.06	0.23	0.01
Swedish Party	10632	0.06	0.24	14888	0.05	0.22	0.01
Christian Party	10632	0.01	0.10	14888	0.01	0.10	0.00
Other parties	10632	0.02	0.15	14888	0.02	0.15	0.00

Bandwidth = 1.00	Elected (N = 14138)			Not elected (N = 23320)			
Variable	N	Mean	Std. dev.	N	Mean	Std. dev.	Difference
Coalition Party	14138	0.17	0.38	23320	0.21	0.41	-0.04
Social Democrats	14138	0.22	0.42	23320	0.29	0.45	-0.06*
Center Party	14138	0.42	0.49	23320	0.33	0.47	0.09*
True Finns	14138	0.01	0.09	23320	0.01	0.09	0.00
Green Party	14138	0.02	0.14	23320	0.03	0.18	-0.01
Socialist Party	14138	0.06	0.24	23320	0.06	0.23	0.01
Swedish Party	14138	0.06	0.24	23320	0.05	0.22	0.01
Christian Party	14138	0.01	0.11	23320	0.01	0.09	0.00
Other parties	14138	0.02	0.15	23320	0.02	0.14	0.00

Notes: * and ** denote 5 % and 1 % statistical significance of difference in means respectively. The significance of differences is tested using t test adjusted for clustering by municipality. 0.55 bandwidth equals roughly the optimal bandwidth chosen using Imbens and Kalyanaraman's (2008) algorithm. Sample includes only candidates running in 1996-2008 elections. Lotteries have been excluded. In 1996 elections income data are available only for candidates who run also in 2000, 2004 and 2008 elections. Income is expressed in euros.

Table F2. Covariate balance within small bandwidths (municipality characteristics).

Bandwidth = 0.01

Variable	Elected (N = 37) N	Mean	Std. dev.	Not elected (N = 38) N	Mean	Std. dev.	Difference
Total number of votes	37	85715	91807	38	81938	84156	3777
Coalition Party seat share	37	25.45	8.06	38	25.88	7.72	-0.42
Social Democrats seat share	37	24.72	7.91	38	24.62	7.11	0.11
Center Party seat share	37	13.84	17.27	38	12.52	15.01	1.32
True Finns seat share	37	1.68	3.14	38	1.46	2.71	0.21
Green Party seat share	37	10.72	7.85	38	10.48	7.44	0.24
Socialist Party seat share	37	9.68	6.17	38	10.19	6.11	-0.50
Swedish Party seat share	37	6.99	14.64	38	7.41	14.76	-0.41
Christian Party seat share	37	3.47	1.83	38	3.56	1.73	-0.08
Other parties' seat share	37	3.46	4.57	38	3.93	4.64	-0.47
Voter turnout	36	57.33	5.43	37	57.17	5.10	0.16
Population	37	190585	204484	38	182109	190372	8477
Share of 0-14-year-olds	36	16.85	2.72	37	16.73	2.39	0.12
Share of 15-64-year-olds	36	68.61	2.55	37	68.75	2.27	-0.14
Share of over-65-year-olds	36	14.54	3.01	37	14.52	2.68	0.02
Income per capita	36	23360	6682	37	23337	6347	23
Unemployment	37	13.63	7.06	38	13.34	6.58	0.29

Bandwidth = 0.05

Variable	Elected (N = 729) N	Mean	Std. dev.	Not elected (N = 761) N	Mean	Std. dev.	Difference
Total number of votes	729	20538	38928	761	21816	40254	-1278
Coalition Party seat share	729	19.88	10.11	761	20.28	10.02	-0.40
Social Democrats seat share	729	23.57	11.09	761	23.80	10.72	-0.23
Center Party seat share	729	27.84	21.17	761	27.21	20.81	0.63
True Finns seat share	729	1.67	3.13	761	1.70	3.41	-0.03
Green Party seat share	729	4.73	5.32	761	4.90	5.37	-0.16
Socialist Party seat share	729	9.42	7.76	761	9.25	7.46	0.17
Swedish Party seat share	729	5.58	17.11	761	5.61	16.78	-0.04
Christian Party seat share	729	3.72	3.47	761	3.92	3.73	-0.20
Other parties' seat share	729	3.34	6.16	761	3.22	5.97	0.11
Voter turnout	711	60.49	6.20	746	60.27	6.40	0.22
Population	720	44984	87091	750	48014	89200	-3030
Share of 0-14-year-olds	711	18.25	3.38	741	18.13	3.22	0.12
Share of 15-64-year-olds	711	65.48	3.10	741	65.64	3.09	-0.16
Share of over-65-year-olds	711	16.27	4.10	741	16.23	4.12	0.04
Income per capita	711	20848	5875	741	21072	5823	-224
Unemployment	720	14.05	6.10	750	13.79	5.96	0.27

Bandwidth = 0.10

Variable	Elected (N = 1778) N	Mean	Std. dev.	Not elected (N = 1906) N	Mean	Std. dev.	Difference
Total number of votes	1778	17428	36257	1906	18328	37762	-901
Coalition Party seat share	1778	19.47	10.24	1906	19.68	10.22	-0.20
Social Democrats seat share	1778	22.61	10.89	1906	22.80	10.65	-0.19
Center Party seat share	1778	30.78	21.57	1906	30.38	21.31	0.40
True Finns seat share	1778	1.78	3.63	1906	1.86	3.81	-0.08
Green Party seat share	1778	4.14	5.10	1906	4.20	5.20	-0.07
Socialist Party seat share	1778	9.03	7.82	1906	8.84	7.66	0.19
Swedish Party seat share	1778	4.99	15.85	1906	5.03	15.87	-0.04
Christian Party seat share	1778	3.54	3.60	1906	3.70	3.74	-0.16
Other parties' seat share	1778	3.52	6.39	1906	3.43	6.23	0.09
Voter turnout	1747	61.36	6.41	1872	61.38	6.47	-0.02
Population	1765	37731	79644	1885	39837	82836	-2105
Share of 0-14-year-olds	1750	18.34	3.31	1871	18.25	3.25	0.09
Share of 15-64-year-olds	1750	64.99	3.20	1871	65.05	3.22	-0.06
Share of over-65-year-olds	1750	16.67	4.36	1871	16.70	4.38	-0.03
Income per capita	1750	20478	5760	1871	20594	5769	-116
Unemployment	1765	13.91	6.00	1885	13.75	5.94	0.16

Bandwidth = 0.20

Variable	Elected (N = 3508) N	Mean	Std. dev.	Not elected (N = 3891) N	Mean	Std. dev.	Difference
Total number of votes	3508	15371	33726	3891	17537	38594	-2166
Coalition Party seat share	3508	19.01	10.43	3891	19.25	10.42	-0.24
Social Democrats seat share	3508	22.31	10.97	3891	22.64	10.87	-0.33
Center Party seat share	3508	31.80	21.41	3891	30.83	21.29	0.97
True Finns seat share	3508	1.65	3.48	3891	1.73	3.72	-0.08
Green Party seat share	3508	3.79	4.95	3891	3.99	5.13	-0.20
Socialist Party seat share	3508	8.82	7.84	3891	8.88	7.77	-0.06
Swedish Party seat share	3508	5.31	16.44	3891	5.38	16.51	-0.08
Christian Party seat share	3508	3.49	3.78	3891	3.52	3.72	-0.03
Other parties' seat share	3508	3.56	6.61	3891	3.62	6.65	-0.06
Voter turnout	3455	61.93	6.48	3833	61.76	6.47	0.18
Population	3490	33380	74698	3865	38006	84731	-4626
Share of 0-14-year-olds	3462	18.38	3.25	3836	18.28	3.23	0.10
Share of 15-64-year-olds	3462	64.73	3.20	3836	64.89	3.24	-0.16
Share of over-65-year-olds	3462	16.89	4.42	3836	16.84	4.41	0.06
Income per capita	3462	20213	5801	3836	20328	5637	-115
Unemployment	3490	13.98	6.06	3865	13.93	6.03	0.05

Bandwidth = 0.30

Variable	Elected (N = 5172) N	Mean	Std. dev.	Not elected (N = 5879) N	Mean	Std. dev.	Difference
Total number of votes	5172	15217	34682	5879	17398	38557	-2180
Coalition Party seat share	5172	18.89	10.39	5879	19.24	10.40	-0.35
Social Democrats seat share	5172	22.15	10.93	5879	22.59	10.89	-0.44
Center Party seat share	5172	32.38	21.35	5879	31.12	21.28	1.26
True Finns seat share	5172	1.66	3.54	5879	1.72	3.70	-0.06
Green Party seat share	5172	3.69	4.93	5879	3.93	5.09	-0.24
Socialist Party seat share	5172	8.69	7.80	5879	8.79	7.75	-0.11
Swedish Party seat share	5172	5.36	16.74	5879	5.32	16.39	0.04
Christian Party seat share	5172	3.46	3.72	5879	3.49	3.65	-0.03
Other parties' seat share	5172	3.53	6.57	5879	3.64	6.63	-0.11
Voter turnout	5100	62.12	6.53	5799	61.87	6.53	0.25
Population	5145	33041	76681	5842	37734	84822	-4693
Share of 0-14-year-olds	5103	18.34	3.25	5798	18.26	3.22	0.08
Share of 15-64-year-olds	5103	64.64	3.23	5798	64.83	3.28	-0.19
Share of over-65-year-olds	5103	17.02	4.48	5798	16.90	4.45	0.11
Income per capita	5103	20078	5719	5798	20259	5655	-181
Unemployment	5145	13.99	6.05	5842	13.97	6.01	0.02

Bandwidth = 0.40

Variable	Elected (N = 6628) N	Mean	Std. dev.	Not elected (N = 7949) N	Mean	Std. dev.	Difference
Total number of votes	6628	15007	35009	7949	18142	40267	-3135
Coalition Party seat share	6628	18.62	10.40	7949	19.30	10.37	-0.68
Social Democrats seat share	6628	21.89	10.94	7949	22.74	10.91	-0.84
Center Party seat share	6628	33.09	21.44	7949	31.06	21.28	2.03
True Finns seat share	6628	1.68	3.60	7949	1.70	3.68	-0.02
Green Party seat share	6628	3.60	4.92	7949	4.01	5.19	-0.41
Socialist Party seat share	6628	8.68	7.81	7949	8.80	7.69	-0.12
Swedish Party seat share	6628	5.28	16.68	7949	5.10	15.90	0.18
Christian Party seat share	6628	3.43	3.71	7949	3.51	3.64	-0.07
Other parties' seat share	6628	3.53	6.67	7949	3.64	6.69	-0.11
Voter turnout	6536	62.26	6.54	7836	61.82	6.51	0.44
Population	6597	32631	77601	7899	39354	88720	-6722
Share of 0-14-year-olds	6543	18.38	3.27	7838	18.25	3.22	0.12
Share of 15-64-year-olds	6543	64.58	3.24	7838	64.87	3.28	-0.30
Share of over-65-year-olds	6543	17.05	4.47	7838	16.88	4.41	0.17
Income per capita	6543	19989	5725	7838	20310	5702	-320
Unemployment	6597	14.05	6.12	7899	13.96	6.00	0.08

Bandwidth = 0.55

Variable	Elected (N = 8710) N	Mean	Std. dev.	Not elected (N = 11348) N	Mean	Std. dev.	Difference
Total number of votes	8710	14284	33359	11348	19657	43620	-5373
Coalition Party seat share	8710	18.43	10.47	11348	19.54	10.39	-1.11
Social Democrats seat share	8710	21.76	11.00	11348	22.90	10.90	-1.13
Center Party seat share	8710	33.42	21.52	11348	30.68	21.35	2.74
True Finns seat share	8710	1.68	3.70	11348	1.70	3.67	-0.02
Green Party seat share	8710	3.51	4.82	11348	4.19	5.39	-0.68
Socialist Party seat share	8710	8.67	7.79	11348	8.82	7.60	-0.14
Swedish Party seat share	8710	5.45	17.10	11348	4.95	15.36	0.49
Christian Party seat share	8710	3.37	3.71	11348	3.48	3.62	-0.11
Other parties' seat share	8710	3.53	6.68	11348	3.60	6.62	-0.07
Voter turnout	8590	62.39	6.57	11183	61.76	6.51	0.63
Population	8672	30986	73934	11284	42703	95855	-11717
Share of 0-14-year-olds	8606	18.37	3.30	11191	18.20	3.22	0.17
Share of 15-64-year-olds	8606	64.48	3.25	11191	64.94	3.34	-0.46
Share of over-65-year-olds	8606	17.14	4.52	11191	16.86	4.45	0.28
Income per capita	8606	19941	5669	11191	20339	5731	-398
Unemployment	8672	14.03	6.09	11284	13.96	6.02	0.06

Bandwidth = 0.70

Variable	Elected (N = 10632) N	Mean	Std. dev.	Not elected (N = 14888) N	Mean	Std. dev.	Difference
Total number of votes	10632	13755	32233	14888	22049	48155	-8295
Coalition Party seat share	10632	18.25	10.51	14888	19.75	10.40	-1.50
Social Democrats seat share	10632	21.61	11.04	14888	22.98	10.83	-1.37
Center Party seat share	10632	33.86	21.66	14888	30.06	21.42	3.81
True Finns seat share	10632	1.70	3.76	14888	1.70	3.63	0.00
Green Party seat share	10632	3.42	4.75	14888	4.45	5.61	-1.03
Socialist Party seat share	10632	8.64	7.83	14888	8.89	7.55	-0.24
Swedish Party seat share	10632	5.43	17.13	14888	4.91	15.05	0.52
Christian Party seat share	10632	3.36	3.69	14888	3.48	3.59	-0.11
Other parties' seat share	10632	3.52	6.67	14888	3.62	6.64	-0.10
Voter turnout	10477	62.49	6.57	14669	61.59	6.49	0.90
Population	10588	29837	71416	14805	47990	105723	-18153
Share of 0-14-year-olds	10498	18.38	3.29	14680	18.18	3.22	0.20
Share of 15-64-year-olds	10498	64.42	3.23	14680	65.09	3.37	-0.67
Share of over-65-year-olds	10498	17.20	4.51	14680	16.73	4.42	0.47
Income per capita	10498	19848	5652	14680	20432	5798	-585
Unemployment	10588	14.07	6.09	14805	13.94	6.01	0.13

Bandwidth = 1.00

Variable	Elected (N = 14138) N	Mean	Std. dev.	Not elected (N = 23320) N	Mean	Std. dev.	Difference
Total number of votes	14138	13094	30932	23320	26432	53737	-13338*
Coalition Party seat share	14138	18.16	10.48	23320	20.46	10.25	-2.31
Social Democrats seat share	14138	21.46	11.11	23320	23.39	10.63	-1.93
Center Party seat share	14138	34.37	21.75	23320	28.52	21.24	5.85
True Finns seat share	14138	1.72	3.80	23320	1.69	3.55	0.03
Green Party seat share	14138	3.28	4.66	23320	4.99	5.97	-1.72*
Socialist Party seat share	14138	8.57	7.80	23320	8.94	7.47	-0.38
Swedish Party seat share	14138	5.42	17.23	23320	4.79	14.45	0.63
Christian Party seat share	14138	3.36	3.73	23320	3.47	3.41	-0.11
Other parties' seat share	14138	3.48	6.84	23320	3.60	6.56	-0.12
Voter turnout	13937	62.66	6.59	22978	61.12	6.41	1.54
Population	14075	28404	68622	23190	57594	118321	-29190*
Share of 0-14-year-olds	13966	18.39	3.29	23001	18.08	3.18	0.31
Share of 15-64-year-olds	13966	64.32	3.22	23001	65.40	3.43	-1.08*
Share of over-65-year-olds	13966	17.29	4.51	23001	16.51	4.36	0.77
Income per capita	13966	19758	5631	23001	20716	5904	-958
Unemployment	14075	14.08	6.09	23190	13.82	5.98	0.26

Notes: * and ** denote 5 % and 1 % statistical significance of difference in means respectively. The significance of differences is tested using t test adjusted for clustering by municipality. 0.55 bandwidth equals roughly the optimal bandwidth chosen using Imbens and Kalyanaraman's (2008) algorithm. Sample includes only data for 1996-2008 elections. Income per capita is expressed in euros.

When is RDD as good as randomly assigned?

The recent literature emphasizes that the local randomization assumption is distinct from the key RDD assumption of no discontinuity in the conditional expectation function of potential outcome. The local randomization assumption is more stringent and not required for RDD. Which of these assumptions is invoked has implications on how to estimate the treatment effect of interest and how to test for the validity of the design (see e.g. de la Cuesta and Imai 2016).

Inspired by the approach proposed by Cattaneo et al. (2015), 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. To determine the largest bandwidth in which the as-good-as-random assumption holds, we either look at the most important covariate or the minimum *p*-value among all the covariates. According to Eggers et al. (2015), incumbency status (elected at t–1) is a reasonable measure of candidate quality. If we use it, bandwidths 0.04 or smaller are as-good-as-random at the 5% significance level (923 non-experimental observations; see Table F1 above). Based on the minimum *p*-value among all the covariates (but not correcting for multiple testing), it seems that bandwidths 0.02 or smaller would be as-good-as random at the 5% significance level (128 observations; again see Table F1 above). These findings indicate that the approach proposed by Cattaneo et al. (2015) leads to rather conservative (small) samples in light of our other RDD findings. This is partly due to not correcting for multiple testing and partly due to the fact that in our election data, many covariates have rather steep slopes with respect to the forcing variable.

It seems that the approach proposed by Cattaneo et al. (2015) is able to reproduce the experimental estimate: When we use these conservative bandwidths, there is no statistically significant difference in the means of getting elected at t+1 elections around the cutoff: The difference is 0.010 (*p*-value 0.32) for the bandwidth of 0.04 and 0.064 (*p*-value 0.75) for the bandwidth of 0.02. However, the smaller bandwidth of 0.02 results in a sample too small to be informative. In that case, the insignificance result arises from the large standard error rather than from a smaller point-estimate. Note that we do not resort here to the randomization inference method proposed by Cattaneo et al. (2015), because we have quite a lot of observations within the two as-good-as-random bandwidths that we consider (see Cattaneo et al. 2016 for a Stata implementation of the randomization inference method).

References

Calonico, S., M. D. Cattaneo, and M. F. Farrell. 2016a. “On the Effect of Bias Estimation on Coverage Accuracy in Non-Parametric Inference.” Forthcoming in *Journal of the American Statistical Association.*

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015. “Optimal Data-Driven Regression Discontinuity Plots.” *Journal of the American Statistical Association* 110(512).

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014a. “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs.” *Econometrica* 82(6): 2295-326.

Calonico, S., M. D. Cattaneo, and R. Titiunik. 2014b. “Robust Data-Driven Inference in the Regression Discontinuity Design.” *Stata Journal* 14(4): 909-46.

Calonico, S., M. D. Cattaneo, M. F. Farrell, and R. Titiunik. 2016b. “rdrobust: software for regressions discontinuity designs.” *Stata Journal* 17 (2): 372-404.

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.

Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare. 2016. “Inference in regression discontinuity designs under local randomization.” *Stata Journal* 16 (2), 331-367.

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.

De Magalhaes, L. 2014. “Incumbency Effects in a Comparative Perspective: Evidence from Brazilian Mayoral Elections.” *Political Analysis* 23(1): 113-26.

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): 259-74.

Imbens, G. and K. Kalyanaraman (2012), “Optimal bandwidth choice for the regression discontinuity estimator.” *Review of Economic Studies* 79 (3), 933-959.

Lee, D. S. 2008. “Randomized experiments from non-random selection in U.S. House elections.” *Journal of* Econometrics 142(2): 675-97.

McCrary, J. 2008. “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test.” *Journal of Econometrics* 142(2): 698-714.

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): 169-86.

Uppal, Y. 2010. “Estimating Incumbency Effects In U.S. State Legislatures: A Quasi-Experimental Study.” *Economics and Politics* 22(2): 180-99.