Assessing the External Validity of Election RD Estimates: An Investigation of the Incumbency Advantage

Author(s): Jens Hainmueller, Andrew B. Hall, and James M. SnyderJr.
Source: *The Journal of Politics*, Vol. 77, No. 3 (July 2015), pp. 707-720
Published by: The University of Chicago Press on behalf of the Southern Political Science Association
Stable URL: http://www.jstor.org/stable/10.1086/681238
Accessed: 04/12/2015 20:25

Assessing the External Validity of Election RD Estimates: An Investigation of the Incumbency Advantage

Jens Hainmueller, Stanford University
Andrew B. Hall, Harvard University
James M. Snyder Jr., Harvard University

The regression discontinuity (RD) design is popular because it provides a design-based estimate of the incumbency advantage. However, the RD estimate is "local": it only identifies the effect in hypothetical elections with a 50-50 tie between the Democratic and Republican candidates. There is significant uncertainty and disagreement over the incumbency effect in safer districts away from this threshold. Indeed, mirroring the competing arguments in the theoretical literature, a survey of political scientists reveals that roughly equal numbers of respondents predict the effect to be either larger, smaller, or the same in less competitive districts. We employ a new method based on a validated conditional independence assumption that allows us to estimate the effect of incumbency in districts in a window around the threshold as large as 15 percentage points—that is, elections in which the winning candidate secured as much as 57.5% of the two-party vote. We find that the incumbency advantage is no larger or smaller in these less competitive cases.

A large body of research is devoted to measuring the electoral advantage (or disadvantage) that incumbent parties and candidates possess purely by virtue of holding office. Researchers care about this quantity because it sheds light on the behavior of voters, the incentives that reelection-motivated legislators face, the hurdles confronting potential opposition candidates, the likelihood that electoral outcomes will respond to shifts in voters' preferences, and other important features of the electoral system. Measuring the incumbency advantage is difficult, however, due to well-known problems of selection and unobserved heterogeneity that bias simple comparisons between incumbent and challenger electoral outcomes. In recent years, the regression discontinuity (RD) design has become a popular method for estimating the incumbency advantage (Lee 2008), because it solves these problems by focusing on close elections where the winner is "as-if" randomly assigned, providing design-based estimation of causal effects with few assumptions.

One significant cost to obtaining this as-if-random variation, however, is that the RD estimator is *local*. It only estimates incumbency advantages for extremely close elections—technically, in fact, it only provides an estimate for hypothetical districts with exactly tied elections.[1] We would like to know about incumbency effects in districts away from the 50% threshold for at least two reasons. First, as we show below, these districts comprise a much larger share of all races. An estimator based on these districts is therefore more informative about US elections in general and far more externally valid. Second, examining the variance in the incumbency advantage across these competitive districts informs our theories about incumbents and US elections. In particular, it helps to resolve the considerable disagreement

Jens Hainmueller (jhain@stanford.edu, http://www.stanford.edu/~jhain/) is an associate professor, Department of Political Science and Graduate School of Business, Stanford University. Andrew B. Hall (hall@fas.harvard.edu, andrewbenjaminhall.com) is a PhD candidate at Harvard University. James M. Snyder Jr. (jsnyder@gov.harvard.edu, http://scholar.harvard.edu/jsnyder/home) is Leroy B. Williams Professor of History and Political Science at Harvard University.

Data and supporting materials necessary to reproduce the numerical results in the article are available in the *JOP* Dataverse (https://dataverse.harvard.edu/dataverse/jop). An online appendix containing supplemental analyses is available at http://dx.doi.org/10.1086/681238.

1. Alternative approaches to provide estimates in these less competitive districts must rely on stronger assumptions to obtain variation in incumbency—e.g., no strategic retirements or exogenous switches in party control conditional on a set of control variables.

The Journal of Politics, volume 77, number 3. Published online April 29, 2015. http://dx.doi.org/10.1086/681238

that exists both in the theoretical literature and the field about the relationship between the incumbency effect and district competitiveness. We discuss both of these in more detail below.

In this article, we employ a new technique—developed in Angrist and Rokkanen (forthcoming)—that allows researchers to generalize the RD estimate beyond 50-50 districts.[2] The method relies upon identifying a set of control variables that constitute a kind of sufficient statistic for the "forcing variable" in the RD in a window around the discontinuity threshold at 50-50. The key insight is that, unlike the usual control strategies, in the context of the RD the conditional independence assumption offers explicitly testable implications. We apply the method to elections for US statewide offices over the period 1946–2012. Using the validity tests, we first show that we can obtain valid estimates for elections in a window as large as 15 percentage points around the threshold—that is, for all elections where the Democratic candidate received between 42.5% and 57.5% of the two-party vote (and also for smaller windows). We then show, using a variety of control strategies including regression and several types of matching, that the average incumbency advantage for these cases is no different than the RD estimate of the incumbency advantage at the threshold. Because any remaining bias in the control strategy should bias us toward finding a growing incumbency advantage, this is strong evidence that the advantage is no larger in these cases. Finally, we employ several descriptive analyses to investigate why the incumbency advantage appears flat across these district types. Incumbents in these less competitive districts continue to exert equal amounts of campaign effort and are no more or less able to scare off challengers with previous office-holder experience in these races.

The article is organized as follows. First, we motivate our study in several ways. We discuss theoretical reasons to expect a larger, smaller, or equal effect away from the RD threshold, we present the results of a survey of political scientists that shows widespread disagreement over whether the effect ought to be larger or smaller away from the threshold, and we present descriptive evidence that obtaining an estimate even in relatively small windows around the RD threshold can markedly increase the estimate's coverage and thus its pertinence. Second, we lay out the technical details of the method. Third, we apply the method to US statewide offices, presenting the results of the validity tests and the estimates of the incumbency advantage away from the threshold. Fourth, we briefly consider evidence to help explain our findings, focusing on incumbent effort and scare-off effects away from the RD threshold. Finally, we conclude.

MOTIVATION

The RD estimate only measures incumbency advantages in 50-50 elections

A growing literature in political science applies the regression discontinuity design to elections to study the incumbency advantage.[3] The logic of the RD is that extremely close elections are, in the limit, "as good as" random, with incumbency status randomly given to either the Democrat or Republican candidate. In the limit—at the threshold when elections are exactly tied, hypothetically—the causal effect of party incumbency status is cleanly identified with weak assumptions.[4]

In order to obtain this clean identification, the RD must estimate only a local average treatment effect (LATE), which is directly valid only for 50-50 elections and perhaps applies to the types of districts and years that experience very close elections. Although the need for unbiased estimates exceeds the gain from reporting biased estimates from larger samples—"better LATE than nothing" as Imbens (2010) writes—the majority of US elections are not extremely close (see table 2), and a more generalizable estimate, if also unbiased, would be valuable. It would immediately speak to the effect of incumbency in a much wider set of districts, and it could test theories of incumbency that predict varying effects across district types.

Theories predict varying incumbency advantages across district types

There are competing arguments in the literature regarding the relationship between the incumbency advantage and district competitiveness. Some predict that the incumbency advantage will be larger in competitive districts than in safe districts. Incumbents in competitive districts may exert more effort

2. Angrist and Rokkanen (forthcoming) apply the technique to an RD involving test scores and admission to exam schools in Boston.

3. This literature includes Boas and Hidalgo (2011), Broockman (2009), Butler (2009), Cellini, Ferreira, and Rothstein (2010), Dal Bó, Dal Bó, and Snyder (2009), DiNardo and Lee (2004), Eggers and Hainmueller (2009), Ferreira and Gyourko (2009), Folke and Snyder (2012), Fouirnaies and Hall (2014), Gagliarducci and Paserman (2012), Gerber and Hopkins (2011), Hainmueller and Kern (2008), Lee, Moretti, and Butler (2004), Leigh (2008), Pettersson-Lidbom (2008), Trounstine (2011), and Uppal (2009, 2010).

4. For an overview of the assumptions, see Lee (2008) as well as Imbens and Lemieux (2008). For empirical challenges to these assumptions in the case of the US House, see Caughey and Sekhon (2011), Grimmer et al. (2012), and Snyder (2005), but also see Eggers et al. (2015) for evidence that the assumption is widely plausible and that the US House may be an exception and not evidence of sorting.

to utilize the direct office holder benefits since they are more vulnerable (e.g., Ashworth and Bueno de Mesquita 2006; Oppenheimer 2005; Stein and Bickers 1997).[5] Incumbents who won by narrow margins in their previous races might also feel vulnerable, leading them to work harder to win reelection. Another hypothesis is that incumbent quality might be higher in competitive districts since these districts are better at weeding out weak incumbents (Erikson 1971).[6,7]

On the other hand, it is also possible that the incumbency advantage is larger in safer districts or after large victories. This would be true if, for example, the ability of incumbents to scare off high-quality challengers is an important component of the incumbency advantage, and if there is an interaction between district safety and scare-off.[8] Suppose, for example, that direct officeholder benefits increase the expected vote for an incumbent by 5 percentage points. Then the scare-off effect might be much smaller in a district with a 50-50 election than in a district that already leans toward the incumbent's party by, say 55%–45%. In the former case, the officeholder benefits move the expected vote for the challenger's party to 45%. This race is leaning toward the incumbent but still potentially winnable, so the challenging party may be able to find a high-quality candidate. In the latter case, however, the incumbency advantage moves the expected vote for the challenger's party to 40%. This puts the race more in the "hopeless" category, making it more difficult for the challenging party to find a high-quality candidate.

Finally, a common assumption in the literature is that all elected officials are constantly worried about losing the next election—that is, they are all "running scared" (e.g., King 1997; Mann 1977). If so, and if the incumbency advantage is mainly due to officeholder benefits, then we might predict that the incumbency advantage will be approximately the same across all types of districts.

Empirically, Hirano and Snyder Jr. (2009) find evidence that state legislative incumbents have a smaller overall electoral advantage in safe districts, and that incumbents' direct office holder benefits are also smaller in safe districts.[9] Desposato and Petrocik (2003) study US House members and find evidence that the incumbency advantage is larger in areas that are "unfriendly" to the incumbent in terms of partisan affiliation, that is, in areas where more voters are registered with the opposition party.

No consensus about incumbency advantage in less competitive districts

Although the theoretical literature does not provide clear predictions about how the incumbency advantage varies across more and less competitive districts, it is nonetheless possible that there is a clear consensus among scholars about the empirical patterns. To get a sense about whether or not this is the case, we administered a short online survey. The bottom line is that there does not appear to be a consensus at all.

The survey focused on statewide elections in the United States and first informed respondents that the party incumbency effect at the 50-50 threshold (based on RD estimates) is about an 8–9 percentage point gain in two-party vote share for the incumbent party. We then asked respondents whether they expect the party incumbency effect in districts where the winner received between 50% and 60% of the vote to be smaller or larger than in districts right at the 50% threshold. Answer options included: larger, about the same, and smaller. If respondents answered larger or smaller, we also asked them about the expected magnitude of the effect.[10] We recruited respondents through political science e-mail lists such as PolMeth (the e-mail list of the Society of Political Methodology). Overall, 165 respondents answered the survey. About 41% indicated that they work in the field of American politics, 25% said they work in comparative politics, and another 25% said they work in methodology.

Table 1 reports the survey results. There is no consensus about the incumbency effect in less competitive districts. About 36% of respondents think that the incumbency effect in less competitive districts is larger than at the threshold. In contrast, another 31% of respondents think that the in-

5. Oppenheimer (2005, 148) writes, "Personal incumbency advantage is reaped by those who need it but not by those who do not. With fewer members needing personal incumbency advantage in the 1990s, it has declined."

6. Erikson (1971, 396) writes, "a Congressman who consistently wins in a district where his party is weak may owe his incumbency to the fact that he is a strong candidate rather than owe his victories to the fact that he is an incumbent." Also see Zaller (1998).

7. In the case of extremely partisan districts or areas there might even be a purely "mechanical" effect reducing the incumbency advantage: if, say, 90% of the voters already identify with the incumbent's party (as in some majority-minority districts), then there are few opposition or independent voters for the incumbent to win over, so the incumbency advantage almost has to be smaller.

8. See, e.g., Banks and Kiewiet (1989), Canon (1993), Cox and Katz (1996), Jacobson and Kernell (1983), Kazee (1994), and Maisel and Stone (1997) for arguments and evidence regarding strategic candidate entry and scare-off. Canon (1993, 1134–35) notes that, "Ambitious amateurs are more likely to run when incumbents are more vulnerable (as indicated by scandal, a strong challenge in the primary, or relatively low vote in the previous election), or when the challenger's party's normal vote is high."

9. They also find some evidence that the incumbent "quality advantage" is larger in competitive districts.

10. The question wording is provided in the appendix.

Table 1. Expectations about the Party Incumbency Effect in Less Competitive Districts Where the Winner Received between 50% and 60% of the Vote

	Larger	Smaller	Same
Do you expect the incumbency effect in these districts to be smaller or larger than at the 50% threshold?			
% of respondents	35.8	30.9	33.3
What effect magnitude do you expect in these districts?			
25th percentile	9.8%	2.5%	
Mean	11.3%	3.4%	
Median	11.3%	4.5%	
75th percentile	12.7%	5.6%	

Note. Results from survey of 165 respondents recruited via Political Science e-mail lists.

cumbency effect is smaller, and about 33% of respondents think that the effect is about the same as in 50-50 elections. Among those who think that the incumbency advantage is larger, the average expected effect magnitude is 11.3 percentage points, about 32% higher than the RD estimate for the incumbency effect at the threshold. Among those who think that the incumbency advantage is smaller, the average expected effect magnitude is 3.4 percentage points, about 60% lower than the effect at the threshold. As a robustness check we also replicated the results only for those respondents that said that they work in American politics; the results are almost identical.

There seems to be considerable disagreement among scholars about the incumbency effect in less competitive districts, a fact that is roughly consistent with the ambiguous theoretical expectations outlined above.

Generalizing the RD estimate would cover a wider set of electoral settings

Regardless of the exact window size and specification used, RDs only provide unbiased estimates of treatment effects at the 50-50 threshold. It is instructive nonetheless to examine salient characteristics of the samples in various windows, to get some sense of the external validity problem inherent in RD analyses. Examining salient characteristics of the samples in various windows is also essential for understanding the relative benefits of being able to move away from the threshold.

In table 2 we present a few summary statistics for different window sizes, for statewide elections in the United States over our period of from 1946 to 2012. The top panel (first three rows) presents three measures of coverage. Notice first that even windows that are relatively wide from an RD point of view—such as the window defined by a 2 percentage-point vote margin—contain only about 8% of the races and 9% of the US population. To cover more than 50% of the races and population, one must consider margins closer to 15 percentage points, which are far outside the typical RD windows. The third row of the top panel shows that even if we consider a generous definition of coverage—what fraction of the states in any given decade have at least one race inside the window—less than half of the state-decades are included in the 1% window, and less than two-thirds are included in the 2% window. On the other hand, more than 90% of the state-decades are covered by the 10% window.

The bottom panel in table 2 shows the mean values of several variables—variables that are of interest in studies of

Table 2. Coverage and Representativeness of Various Windows

	Window Size							
	.5%	1%	2%	5%	10%	15%	20%	100%
Percent of races	1.9	4.0	8.2	20.0	36.9	51.7	62.6	100.0
Percent of population	2.1	4.1	8.8	21.7	40.1	56.6	67.8	100.0
% of state-decades with 1+ obs.	23.7	43.7	64.1	86.1	92.2	94.9	95.6	100.0
Percent open seats	37.0	39.8	40.8	46.4	44.8	44.5	42.9	36.4
Average winner vote margin	.26	.51	1.01	2.43	4.75	6.92	8.72	23.01
Average |normal vote—50%|	5.28	5.75	5.04	4.98	5.12	5.37	5.66	9.40
Average |dem. wins—50%|	20.4	21.0	19.8	21.0	21.7	22.4	23.0	26.5
Percent of cases in south	10.9	12.8	10.7	10.4	12.0	12.4	12.4	19.9

Note. Row 3 gives percent of state-years with at least one race in the given window.

the incumbency advantage—for the races that fall inside the various windows. Comparing the numbers in the first seven columns to those in the last column provides some sense of how (un)representative the races inside each window are, at least on average, compared to the sample of all races. Note first that (by definition), in the smaller windows the average vote margin of the winner is much smaller than it is in the full sample of all races. Somewhat more surprisingly, the percentage of races that involve open seats does not change much as the window size increases out to 20%, and the percentage is also not monotonic. The same is true for "overall partisan bias" toward one party, measures by the average deviation between the normal vote and 50%.

The south is rather heavily under-represented in the smaller windows. The smaller windows also include a disproportionate number of races from earlier years but not by too much.

METHOD

In this section we explain the method from Angrist and Rokkanen (forthcoming) in the context of elections, and we describe the estimation strategies we use to implement the method using election data on US statewide offices.

The outcome, $Y_{i,t+1}$, is the Democratic (two-party) vote share in district i in the election in period $t+1$. Let $V_{i,t}$ be the forcing variable, the Democratic vote share winning margin in the district in the previous election in period t. The treatment variable of interest is $D_{i,t} \equiv 1\{V_{i,t} > 0\}$, an indicator of Democratic victory in period t. Accordingly, $Y_{i,t+1}(D_{i,t})$ are the two potential outcomes of interest that capture the vote shares that the Democrats attain in a district at $t+1$ if they win or loose the election in the district at t. The average difference between these two potential outcomes is the party incumbency effect.

As Angrist and Rokkanen (forthcoming) point out, $D_{i,t}$ is a deterministic function of the forcing variable and, therefore, in comparing treated and control units, $V_{i,t}$ is the *only* omitted variable. Put another way, if the forcing variable in any RD is randomly assigned, then the treatment is also randomly assigned, and we could analyze it as an experiment without worrying about the discontinuity threshold or about modeling the forcing variable.

As is well known, the usual RD only identifies the effect of interest at the threshold where, in the limit, treated and control units have the same value of the forcing variable. Away from the threshold there is no overlap in the forcing variable, as all elections with $V_{i,t} > 0$ are treated with a Democratic incumbent and all elections with $V_{i,t} < 0$ are treated with a Republican incumbent.

To get away from the discontinuity threshold, Angrist and Rokkanen (forthcoming) propose gathering a set of control variables, $X_{i,t}$, and imposing a conditional independence assumption (CIA), which asserts that

$$E[Y_{i,t+1}(D_{i,t})|V_{i,t},\ X_{i,t}] = E[Y_{i,t+1}(D_{i,t})|X_{i,t}]$$

for $D_{i,t} \in \{0, 1\}$. This says that once we condition on the set of covariates in $X_{i,t}$, the potential outcomes are mean-independent of the forcing variable $V_{i,t}$. In other words, by controlling for the set of covariates we break the correlation between the forcing variable and the potential outcomes, ensuring that we can identify the missing counterfactual average of what would have happened to the treated units in the absence of the treatment. In particular, if the CIA holds then in this conditioning set we have that $(Y_{i,t+1}(1), Y_{i,t+1}(0)) \perp V_{i,t}|X_{i,t}$ and, given common support, the average treatment effect on the treated (ATT)[11] is therefore identified by a covariate adjusted comparison of the observed outcomes for treated and control units, as in

$$\begin{aligned}\tau_{ATT} &= E[Y_{i,,t+1}(1) - Y_{i,t+1}(0)|D_{i,t} = 1]\\ &= \int (E[Y_{i,,t+1}|D_{i,t} = 1,\ X_{i,t}] - E[Y_{i,t+1}|D_{i,t} = 0,\ X_{i,t}])dP(X|D_{i,t} = 1).\end{aligned}$$

Finding such a conditioning may prove difficult in many settings. But the RD setup has a unique advantage in that it generates a clear, testable implication for the CIA. Instead of simply asserting that a particular set of control variables are sufficient to produce causal estimates, the RD setup allows us to test whether the assumption appears plausible—by which we mean to see whether we detect evidence that it is violated.

Consider races to the right of the threshold where $D_{i,t} = 1$. The CIA assumption implies that

$$\begin{aligned}E[Y_{i,t+1}(1)|V_{i,t},\ X_{i,t},\ D_{i,t} = 1] &= E[Y_{i,t+1}(1)|X_{i,t}]\\ &= E[Y_{i,t+1}(1)|X_{i,t},\ D_{i,t} = 1],\end{aligned} \tag{1}$$

which means that we should see that

$$E[Y_{i,t+1}|V_{i,t},\ X_{i,t},\ D_{i,t} = 1] = E[Y_{i,t+1}|X_{i,t},\ D_{i,t} = 1] \tag{2}$$

11. Our focus on the ATT is arbitrary since in the RD the party of interest and the definition of the "treatment" is irrelevant due to symmetry. We could also estimate the ATC, where we match Democratic-winner units to Democratic-loser units. Equivalently, we could redefine the treatment to be Republican victory. The ATT from the Republican-treatment definition would be the same as the ATC in our given setup and vice versa. In the appendix, we estimate this alternative quantity. As expected, we find highly similar results. See figure A.2 and surrounding discussion.

if the proposed conditioning set $X_{i,t}$ makes the CIA valid. By the same logic, for races to the left of the threshold we should see

$$E[Y_{i,t+1}|V_{i,t},\ X_{i,t},\ D_{i,t}=0]=E[Y_{i,t+1}|X_{i,t},\ D_{i,t}=0]. \tag{3}$$

In practice, we can test for this by estimating OLS equations of the form

$$Y_{i,t+1}=\beta_0+\beta_1 V_{i,t}+X_{i,t}+\varepsilon_{i,t} \tag{4}$$

on each side of the threshold (in separate regressions) and testing for $\beta_1=0$. As Angrist and Rokkanen (forthcoming) point out, it is likely that the CIA only holds within some window around the threshold.[12]

The intuition of the tests is as follows. The RD tells us that $V_{i,t}$ is the only omitted variable. If our control set successfully addresses the omitted variable bias, then when it and $V_{i,t}$ are included in the same regression on each side of the discontinuity, there should be no remaining correlation between $V_{i,t}$ and the outcome variable. If we find that there is no remaining correlation—in other words, that β_1 is close to 0—then we have reason to believe we can compare treated and control units with like values of X even though they have different values of the forcing variable. This allows us to move away from the threshold even though by definition no treated and control units have overlapping values of the forcing variable.

Because failing to reject the null hypothesis that $\beta_1=0$ constitutes a "pass" for the CIA's validity test, it is important to consider issues of power. Most importantly, a failure to reject the null should only be considered evidence for the CIA if β_1 is substantively small. A large but noisy estimate of β_1 is a red flag. To make these ideas especially clear, in the online appendix we offer a counterexample using the US House where the CIA tests do not appear to hold.

To estimate the incumbency advantage under this CIA assumption, we employ three different methods for the covariate adjustment to ensure that the results are robust. First, we use a simple OLS regression where the outcome $Y_{i,t+1}$ is regressed on the treatment indicator $D_{i,t}$ and the control variables included in the conditioning set $X_{i,t}$. Second, to relax the linearity assumption we use one-to-one genetic matching with bias adjustment where we match each treated unit to the closest control unit based on a generalized Mahalanobis distance (Diamond and Sekhon 2013).[13] Third, we employ an entropy balancing where we impose exact balance on the first moments of each of the control variables in $X_{i,t}$ (Hainmueller 2012). For each effect estimation the sample is restricted to a window around the threshold where the CIA appears to be valid. We also assess the overlap in the adjusted data using standard balance checks (see the appendix for results).

APPLICATION TO US STATEWIDE OFFICES

Data

We apply the method to the data on US elections for statewide offices, 1946–2012. The specific offices we use are: Attorney General, Auditor, Governor, Lieutenant Governor, Senator, Secretary of State, and Treasurer.[14] The data set has been compiled from a variety of state sources. See Ansolabehere and Snyder (2002) for details. For details on the exact numbers of observations for each office and state, along with the range of years covered for each state, see table A.1 in the appendix.

Testing the conditional independence assumption

Table 3 presents the results for the tests of the CIA described in equation (4), where in a given window to the left or the right of the threshold, we regress the Democratic vote share in the election at $t+1$ on the forcing variable, that is, the Democratic winning margin at t, and a set of control variables. The quantity of interest is β_1, the coefficient on the forcing variable. If conditional independence holds conditional on the control variables, then we expect the coefficient on the forcing variable to be close to zero.

We use three different conditioning sets for the control variables. The first set is the most extensive and includes two lags of the Democratic vote share in the $t-1$ and $t-2$ elections, two lags of the normal vote in the $t-1$ and $t-2$ elections, and a variable that measures the "Midterm Slump" at t. The normal vote is calculated as the average of the Democratic vote share across all offices in the state for the given election cycle. The midterm slump variable takes the value 1 for midterm elections after a Democratic presidential victory and -1 for midterm elections after a Republican presidential victory, and 0 otherwise. We choose

12. For convenience, we will focus on symmetric windows around the threshold, but there is no reason we could not use a window that includes a larger region of treated or control cases.

13. The method uses a genetic search algorithm to determine the weight that is given to each control variable in order to maximize covariate balance.

14. In the main analyses we leverage all close elections from these offices. However, in unreported results we have also examined the possibility of heterogeneity in the effects across offices. We find no difference in the effect for "high" offices—senator and governor—vs. other offices.

Table 3. Conditional Independence Tests

Window	Control Set 1: Dem Share$_{t-1}$ Dem Share$_{t-2}$ Normal Vote$_{t-1}$ Normal Vote$_{t-2}$ Midterm Slump$_t$		Control Set 2: Dem Share$_{t-1}$ Dem Share$_{t-2}$ Normal Vote$_{t-1}$ Midterm Slump$_t$		Control Set 3: Dem Share$_{t-1}$ Normal Vote$_{t-1}$	
	$D=0$	$D=1$	$D=0$	$D=1$	$D=0$	$D=1$
5	−.08 (.33)	.32 (.29)	−.07 (.33)	.27 (.30)	.00 (.31)	.21 (.28)
	$N=441$	$N=446$	$N=441$	$N=446$	$N=471$	$N=474$
10	.07 (.11)	.07 (.12)	.07 (.11)	.06 (.12)	.12 (.11)	.09 (.12)
	$N=837$	$N=811$	$N=837$	$N=811$	$N=899$	$N=866$
15	.29 (.07)	.06 (.07)	.29 (.07)	.05 (.07)	.30 (.07)	.11 (.07)
	$N=1170$	$N=1132$	$N=1170$	$N=1132$	$N=1255$	$N=1201$
20	.32 (.06)	.11 (.05)	.32 (.06)	.11 (.05)	.33 (.05)	.15 (.05)
	$N=1389$	$N=1387$	$N=1389$	$N=1387$	$N=1485$	$N=1471$
25	.31 (.04)	.18 (.04)	.31 (.04)	.18 (.04)	.31 (.04)	.23 (.04)
	$N=1553$	$N=1615$	$N=1553$	$N=1615$	$N=1655$	$N=1709$
30	.29 (.04)	.20 (.03)	.29 (.04)	.21 (.03)	.30 (.04)	.25 (.03)
	$N=1655$	$N=1783$	$N=1655$	$N=1783$	$N=1761$	$N=1879$
35	.30 (.04)	.22 (.03)	.30 (.04)	.22 (.03)	.30 (.03)	.26 (.03)
	$N=1736$	$N=1898$	$N=1736$	$N=1898$	$N=1844$	$N=2003$
40	.29 (.03)	.18 (.03)	.29 (.03)	.18 (.03)	.30 (.03)	.22 (.03)
	$N=1783$	$N=1980$	$N=1783$	$N=1980$	$N=1894$	$N=2093$

Note. CIA tests from equation (4) to the left of the discontinuity ($D=0$) and to the right ($D=1$). The CIA appears to be satisfied at windows as large as size 10 and partially satisfied at 15. Robust standard errors in parentheses. V_i,t and $Y_i,t+1$ measured in percentage points.

these covariates because we expect them to be highly predictive of the outcome and correlated with the forcing variable. The second conditioning set omits the midterm slump variable and the third set in addition omits the second period lags for the normal vote and the Democratic vote share.

Table 3 shows the β_1 estimates for windows to the left of the threshold defined by anywhere from a 5% window to a 40% window (incremented by 5%). For example, the 5% window contains elections in which the Democratic percentage of the two-party vote was between 47.5% and 52.5%. To the right of the threshold ($D=1$), this contains districts where the Democratic vote share was between 50% and 52.5% percent and the same window to the left of the threshold ($D=0$) contains districts where the Democratic vote share was between 47.5% and 50%. We find that the β_1 estimates are close to zero and insignificant at conventional levels for all the 5% and 10% windows across all three covariate sets.

To make this point clear, consider the estimates in the first two columns in the second row, for the 10% window. Controlling for the five variables in control set 1, we see that the remaining, conditional relationship between the running variable is 0.07, for both the $D=0$ and $D=1$ sides. This means that a 1 percentage-point increase in the Democratic vote-share winning margin at time t is associated conditionally with a 0.07 percentage-point increase—that is, a 7 basis point increase—in the Democratic vote share at time $t+1$. This is a substantively tiny association: if we were to change the Democratic vote-share winning margin by a full 100 percentage points, we would see only a 7 percentage-point change in the vote share the next time around. This is the basis on which we can say that the estimate is substantively small. We can also say that this estimate is precise. The upper and lower bounds of its 95% confidence interval (which range from roughly −0.15 to 0.29) are still relatively small, meaning that we can reject hypotheses of there being a large conditional link between the two variables. It is these two facts together—the substantively small effect, and the precision with which we estimate it—which lead us to conclude that the test is supportive of the CIA.

Although the $D=1$ estimates for the 5% window are slightly larger in magnitude, between .21% and .32%, the fact that the $D=0$ estimates are quite small for the 5% window

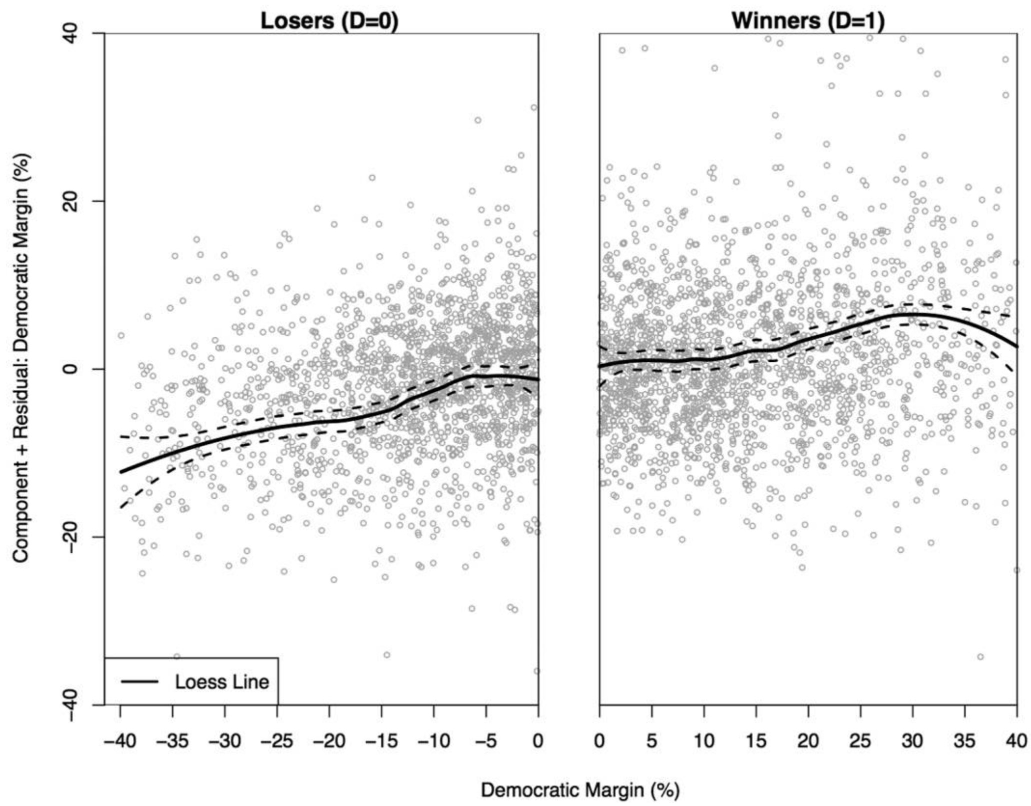

Figure 1. Component residual plot to test the conditional independence assumption. The figure shows component residual plots that summarize the relationship between the forcing variable (Democratic winning margin) and the outcome (Democratic Vote Share in the next election) after partialling out the linear component of the control variables (based on conditioning set 1). Loess lines with 95% pointwise confidence intervals are added to approximate the partial regression function.

(within .08% around zero), as are all estimates for the 10% window, suggests this is the result of sampling variability rather than an underlying violation of the CIA.[15]

For the 15% window—when the Democratic percentage of the two-party vote is between 42.5% and 57.5%—the estimates for the window to the right of the threshold are still close to zero, but for the window to the left of the threshold the estimates are positive and significant, indicating that the conditional independence assumption is beginning to fail as we move further away from the threshold into much less competitive districts. For windows of 20% or larger the estimates are now all positive and significant indicating that the conditional independence assumption is invalid because the forcing variable is informative about the outcomes even conditional on the control variables.

The success of the CIA tests at smaller windows, but not at larger windows, reflects the bounded nature of the assumption. For any election data set, it is likely to hold only for some window around the threshold in which there is enough random variation in the presence of close elections to make some cases where Democratic candidates win comparable to some cases where Democratic candidates lose. Farther from the threshold, districts are systematically partisan and it becomes increasingly difficult for any set of control variables to condition out the extremely strong link between the Democratic win margin at time t and the Democratic vote share at time $t + 1$, both of which are likely to be either quite low or quite high.

In order to visually inspect the conditional independence assumption, figure 1 shows the component residual plots for the regressions to the left and the right of the threshold (based on conditioning set 1). The loess lines summarize the relationship between the forcing variable and the outcome after partialling out the linear component of the control variables. Where the CIA is met, we should see a flat line indicating no residual relationship between the forcing variable and the outcome, once the control variables have been partialled out of each. The plot confirms the results from table 3. The conditional independence assumption holds well to the right of the threshold, as far as the 20% window. Beyond the 20% Democratic win margin, we see the slope

15. Moreover, these larger coefficients are still very small relative to the effects we uncover below.

begin to increase as we move outside the bounds where the CIA appears to be met. To the left of the threshold we see a similar phenomenon, albeit in a somewhat smaller window.[16] The loess line is flat until about −10%, at which point it begins to bend down and the forcing variable becomes predictive of the outcome. The regions around 0 where the loess lines are flat represent the region where the CIA appears to be valid. These are the sets of elections for which we can generalize the RD estimate that applies only at the threshold.

Incumbency effects in less competitive districts

The previous results suggest that the conditional independence assumption is valid for the 5% and 10% window and partially so for the 15% window. Although the windows for which the CIA appears valid may not contain many races from exceedingly "safe" districts, like districts where the Democratic normal vote is above 60% or below 40%, they do contain a large number of elections for which our survey indicated substantial uncertainty over the effect of incumbency. These "less competitive" districts are important because they comprise a much larger portion of all elections than do elections at the 50-50 threshold. The elections in these districts are also important because they are sufficiently far away from 50-50 to inform us about variation in the incumbency advantage across districts.

The incumbency effect for the elections that fall into these windows is identified by a covariate adjusted comparison of winners and losers. For each of these windows, we compute the incumbency effects using the three different covariate adjustments described above. As a benchmark, we also add the RD design estimates of the incumbency effect at the threshold as a comparison. For the RD estimates we use standard local linear regressions where the outcome variable is regressed on the treatment indicator, the forcing variable, and the interaction between the two. We fit these regressions to samples determined by bandwidths of a 1%, 2%, and 5% margin.

Table 4 displays the results. The bottom panel shows the RD estimates at the threshold, which indicate that in 50-50 elections, party incumbency increases vote shares in the next election by about 8–9 percentage points. The top panel shows the estimates of the incumbency effect in less competitive districts based on the CIA. We find that the incumbency effect in less competitive districts is very similar to the RD incumbency effect at the threshold. The results are precisely estimated and robust across the different windows and covariate adjustment methods, with effect estimates in the range of 8–9 percentage points. In the appendix we present a series of balance tests to establish the success of the control strategies at producing comparable treated and control groups. We also present the results of a standard sensitivity analysis (Rosenbaum 2002), which suggests that the results are not sensitive to omitted confounders.

In figure 2 we also plot the incumbency effect estimates (with 95% confidence intervals) for the sample of districts that fall into increasingly larger windows between 1% and 15% (based on the regression adjustment with conditioning set 1) and compare it to the RD-based incumbency effect estimate at the threshold (based on the local linear regression with a 5% bandwidth). The effect estimates in less competitive districts all fall within the confidence intervals of the RD estimate at the threshold.

Looking at the magnitude of the effect—for example, 8.30% using OLS with control set 1 in a 10% window—we see that it is far larger than predicted by most of the 33.3% of survey respondents who thought the effect would be smaller using less competitive districts and far smaller than predicted by most of the 35.8% of respondents who thought the effect would be larger using less competitive districts (see table 4).

Recall from above that for windows around 15% the CIA tests start to fail (at least to the left of the threshold). Given the strong and positive raw correlation between election outcomes across years—reflected also in the fact that when the CIA tests fail the coefficient on the running variable is always positive—and given that the running variable is the only omitted variable, it is highly likely that any biases in the estimates for windows away from the threshold are positive. This would bias us toward finding systematically higher incumbency effects in large windows. Yet we find that the effects are quite similar across window size, despite any potentially remaining omitted variable bias.

External validity of incumbency effects in less competitive districts

The previous section shows that the incumbency effects for somewhat less competitive districts are very similar to RD based estimates right at the threshold. How externally valid are these results? One potential limitation with the previous analysis is that the estimation windows for the incumbency effects include districts that are very close and those further away from the threshold. For example, the 10% window in-

16. As discussed in the Method section, this suggests that we could use an asymmetric window that includes a larger window of observations to the right of the threshold than to the left. Doing so leads to similar results as the symmetric windows we report.

Table 4. Incumbency Effects in Less Competitive Districts and at the Threshold

Incumbency Effect in Less Competitive Districts									
	Control Set 1: Dem Share$_{t-1}$ Dem Share$_{t-2}$ Normal Vote$_{t-1}$ Normal Vote$_{t-2}$ Midterm Slump$_t$			Control Set 2: Dem Share$_{t-1}$ Dem Share$_{t-2}$ Normal Vote$_{t-1}$ Midterm Slump$_t$			Control Set 3: Dem Share$_{t-1}$ Normal Vote$_{t-1}$		
Window	OLS	Match	Weight	OLS	Match	Weight	OLS	Match	Weight
5	7.99 (.62)	7.91 (.75)	8.32 (.65)	8.01 (.63)	7.21 (.72)	8.32 (.65)	7.57 (.60)	7.85 (.70)	7.86 (.63)
	$N = 887$	$N = 887$	$N = 887$	$N = 887$	$N = 887$	$N = 887$	$N = 945$	$N = 945$	$N = 945$
10	8.30 (.47)	7.85 (.60)	8.53 (.52)	8.32 (.47)	8.06 (.61)	8.53 (.52)	8.13 (.45)	8.25 (.58)	8.36 (.48)
	$N = 1648$	$N = 1648$	$N = 1648$	$N = 1648$	$N = 1648$	$N = 1648$	$N = 1765$	$N = 1765$	$N = 1765$
15	9.32 (.42)	9.39 (.59)	9.56 (.48)	9.33 (.42)	9.20 (.60)	9.56 (.48)	9.29 (.40)	9.84 (.57)	9.54 (.43)
	$N = 2302$	$N = 2302$	$N = 2302$	$N = 2302$	$N = 2302$	$N = 2302$	$N = 2456$	$N = 2456$	$N = 2456$

Incumbency Effect at the Threshold (RD estimates)			
Bandwidth	Local Linear	Local Linear	Local Linear
1	9.99 (3.44)	9.99 (3.44)	9.36 (3.22)
	$N = 178$	$N = 178$	$N = 191$
2	8.73 (2.23)	8.73 (2.23)	8.41 (2.08)
	$N = 361$	$N = 361$	$N = 384$
5	7.52 (1.30)	7.52 (1.30)	7.22 (1.22)
	$N = 887$	$N = 887$	$N = 945$

Note. The top panel presents incumbency effect estimates in less competitive districts based on the conditional independence assumption for different windows and covariate adjustment methods. The bottom panel presents comparison of the incumbency effect estimates at the threshold based on a regression discontinuity design for different bandwidths. Covariate adjustments are OLS: Linear regression; Match: One-to-one nearest neighbor matching with replacement and bias adjustment; Weight: Entropy balancing; Local linear: Local linear RD regression. Robust standard errors in parentheses. Window: Sample used to estimate the effect by comparing winners and losers. Bandwidth: Sample used to estimate the RD effect at the threshold. $Y_i,t+1$ measured in percentage points, 0–100.

cludes hypercompetitive districts where the Democrats won with a razor thin margin of just over zero percentage points, as well as relatively safer districts where the Democrats won with a margin of close to 10 percentage points. Since the overall treatment effect is an average of the effects across all these districts, it might mask important heterogeneity in effect as we move further away from the threshold. To examine the external validity we now replicate the analysis but restrict the sample of winners to districts where the Democratic candidate won with margins between 5%–10%, 5%–15%, and 10%–15%, such that the effect estimates are identified solely based on less competitive districts. The control observations are drawn from districts where the Democratic candidate lost with margins anywhere between −10% and 0%.

Table 5 presents the results. We find that the incumbency effect estimates are very similar to the previous estimates. Even when we exclude districts where the Democrats barely won, the estimated magnitudes are still about 8–9 percentage points. These results are precise and robust across all three windows, covariate adjustments, and conditioning sets.[17] These findings suggest that the incumbency effect in less competitive races with, say, a 10% winning margin is indeed very similar to the incumbency effect at the threshold. Taken together, these additional results corroborate the external validity of the previous findings.

17. As in the previous analysis, balance tests for these control strategies are available in the appendix.

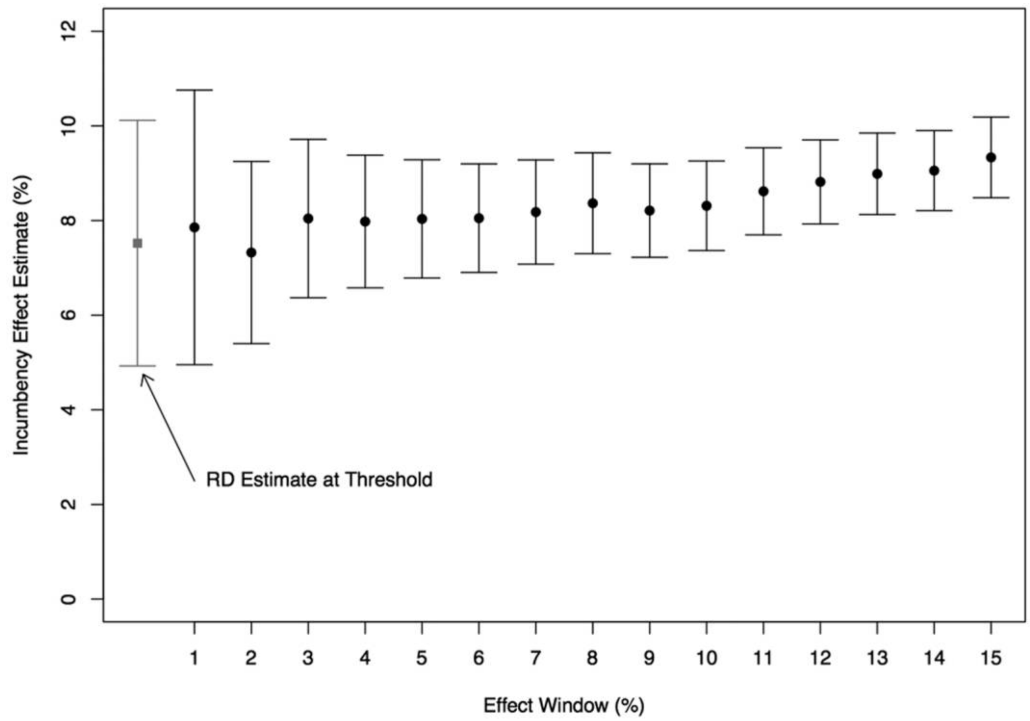

Figure 2. Incumbency effects in less competitive districts and at the threshold. The figure shows the incumbency effect estimates in less competitive districts based on the conditional independence assumption for windows between 1% and 20% (based on the regression adjustment with conditioning set 1). For comparison the figure at the very left also shows the RD based estimate of the incumbency effect at the threshold (based on the local linear regression with a 5% bandwidth).

Table 5. Incumbency Effects in Less Competitive Districts Further Away from the Threshold

	Control Set 1: Dem Share$_{t-1}$ Dem Share$_{t-2}$ Normal Vote$_{t-1}$ Normal Vote$_{t-2}$ Midterm Slump$_t$			Control Set 2: Dem Share$_{t-1}$ Dem Share$_{t-2}$ Normal Vote$_{t-1}$ Midterm Slump$_t$			Control Set 3: Dem Share$_{t-1}$ Normal Vote$_{t-1}$		
Margin	OLS	Match	Weight	OLS	Match	Weight	OLS	Match	Weight
5–10	8.48 (.65)	7.95 (.91)	8.57 (.69)	8.49 (.65)	8.30 (.85)	8.57 (.69)	8.44 (.62)	9.27 (.86)	8.56 (.64)
	$N = 1202$	$N = 1202$	$N = 1202$	$N = 1202$	$N = 1202$	$N = 1202$	$N = 1291$	$N = 1291$	$N = 1291$
5–15	8.78 (.53)	8.67 (.70)	8.83 (.59)	8.78 (.53)	8.44 (.70)	8.84 (.59)	8.89 (.51)	9.37 (.68)	8.98 (.54)
	$N = 1523$	$N = 1523$	$N = 1523$	$N = 1523$	$N = 1523$	$N = 1523$	$N = 1626$	$N = 1626$	$N = 1626$
10–15	9.11 (.69)	9.00 (.93)	9.13 (.75)	9.11 (.69)	8.85 (.91)	9.13 (.75)	9.43 (.67)	9.43 (.85)	9.47 (.69)
	$N = 1158$	$N = 1158$	$N = 1158$	$N = 1158$	$N = 1158$	$N = 1158$	$N = 1234$	$N = 1234$	$N = 1234$

Note. Incumbency effect estimates in less competitive districts based on the conditional independence assumption for different margins and covariate adjustment methods. The estimates are based on districts where the winners won with margins between 5%–10%, 5–15%, or 10–15% and excludes districts where the winner won with narrower margins. Covariate adjustments are OLS: Linear regression; Match: One-to-one nearest neighbor matching with replacement and bias adjustment; Weight: Entropy balancing. Robust standard errors in parentheses. Margins: Sample of winners used to estimate the effect by comparing winners and losers, e.g., margins 5–10 means that only districts where the Democrats won with margins between 5% and 10% are included. For all margins, the sample of losers consists of districts where the Democrats lost with margins between −10 and 0%. $Y_i,t+1$ measured in percentage points, 0–100.

WHY DOESN'T THE INCUMBENCY ADVANTAGE GROW AWAY FROM THE THRESHOLD?

Having laid out our main findings, we now focus on two possible mechanisms to explain why the advantage does not grow away from the threshold. Theories that predict the incumbency advantage to be larger or smaller in less competitive districts often focus on incumbent effort, which might be lower in safer districts, and on the "scare-off" effect, which might be higher in safer districts. Here we offer some evidence that neither of these dynamics changes away from the RD threshold. Incumbents do not appear to exert less effort in these safer races, and they do not appear to scare off challengers with previous officeholder experience at a higher rate.

Incumbent effort constant across district types

To measure incumbent effort, we focus on one observable component: campaign fund-raising. If safer incumbents devote less time to reelection efforts, then we should observe them receiving fewer campaign funds. However, for the window in which we analyze the incumbency advantage, we find an almost perfectly flat relationship between the Democratic vote-share winning margin and total contributions to the Democratic candidate, as figure 3 shows. This descriptive evidence suggests that incumbents' effort does not vary within the window of our analysis.

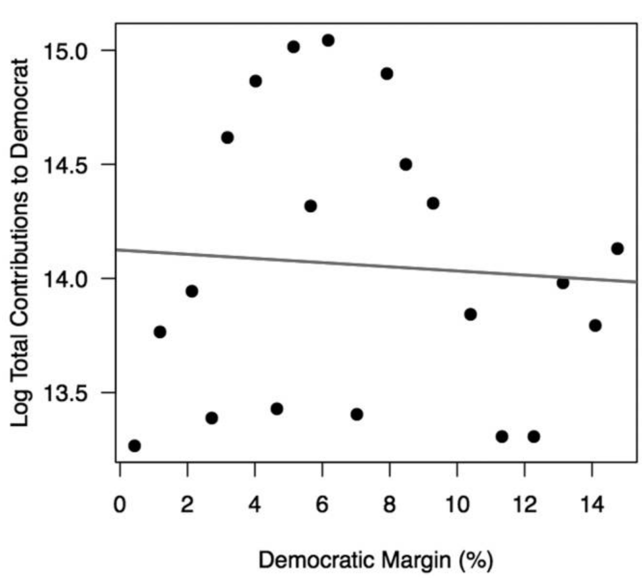

Figure 3. Campaign contributions to the Democratic Party at $t+1$ after a Democratic win at t. The Democrats appear to receive equal levels of contributions in places they barely won and in places they won by as much as 15 percentage points. As a proxy for campaigning effort, the plot therefore suggests that incumbent effort is constant across win margin, helping to explain why we do not find that the incumbency advantage grows (or shrinks) away from the 50-50 threshold. Note: Each point represents an average of the outcome variable in an equal-sample-sized bin of Democratic vote-share winning margin.

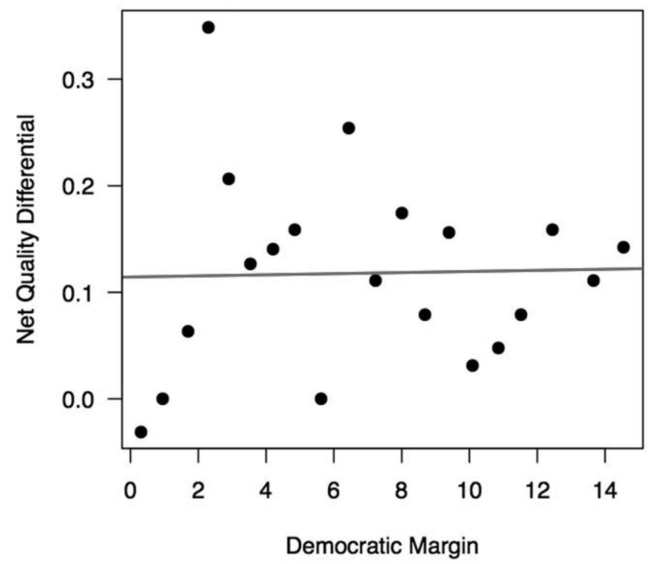

Figure 4. Net quality differential at $t+1$ after a Democratic win at t. The Democrats appear to face equal levels of net quality differential in places they barely won and in places they won by as much as 15 percentage points. As a proxy for scare-off, the plot therefore suggests that the scare-off effect of incumbents is constant across win margin, helping to explain why we do not find that the incumbency advantage grows (or shrinks) away from the 50-50 threshold. Note: Each point represents an average of the outcome variable in an equal-sample-sized bin of Democratic vote-share winning margin.

Scare-off effect constant across district types

In figure 4 we repeat this exercise for the "net candidate quality differential" between Democratic and Republican candidates across the Democratic vote share in the election in order to investigate the scare-off of quality candidates due to incumbency. To measure this "quality" differential, we follow a large previous literature and use an indicator variable for whether each candidate possess previous office-holder experience (e.g., Jacobson 1989). Further details are offered in the appendix. As the figure suggests, scare-off too appears to be flat away from the 50-50 threshold.[18]

CONCLUSION

In this article, we have applied a new technique from the econometrics literature to investigate the causal effect of

18. In the appendix, we extend this analysis by applying the technique to extrapolate the RD effect on scare-off away from 50-50. Here the CIA tests appear valid, and we again find little change in the effect, thus helping to explain why we find a flat electoral incumbency advantage. We do not repeat this exercise for the campaign finance results because the CIA tests do not perform well with that outcome variable.

incumbency on US statewide electoral outcomes away from the 50-50 threshold at which the RD estimate applies. The technique takes advantage of the unique opportunity the RD setup provides to develop a testable, validated Conditional Independence Assumption which is far more plausible than a usual "selection on observables" strategy. We find that the estimated incumbency advantage is just as large when measured in windows as wide as 15 percentage points around the discontinuity threshold. As we showed, estimating the effect in the 15% window includes observations from over 90% of all state-decades and more than half (51.7%) of all races. Thus, while the RD estimate identifies the local average treatment effect only for 50-50 races, it is in fact surprisingly generalizable, at least for statewide US elections.

From a substantive standpoint, the results show that incumbents continue to enjoy the same large electoral advantage in elections away from the threshold, thus including observations in less competitive districts. There does not appear to be any interaction between the incumbency advantage and the previous margin of victory within the window of elections we study. Instead, incumbents continue to enjoy the same advantage, on average, whether they win by 1 percentage point or 15. Even if there is any remaining bias from the validated CIA we employ, this bias would likely lead us to find a growing incumbency advantage. We therefore have strong evidence that the incumbency advantage does not grow for the window we study. This suggests that incumbents do not reduce their efforts to win reelection after winning by relatively wide margins, as some have predicted. Nor are incumbents able to parlay a relatively easy victory in one election into a "free ride" in their next run.

We have also offered some follow-up tests to explain substantively why the incumbency advantage does not change for these races. Even in elections fairly far from 50-50, incumbents continue to raise money at similar rates to those in highly competitive elections. This suggests that incumbents "run scared" and continue to exert effort to campaign in races away from the 50-50 threshold, which helps explain why the advantage remains flat. In addition, incumbents in these elections appear no more or less capable of scaring off potential challengers who possess previous office-holder experience. Within this window of elections, district safety thus does not appear to interact with the strategic decisions of would-be challengers, another reason that the incumbency advantage remains flat.

Finally, the technique we have applied to generalize the RD estimate of the incumbency advantage is likely to be useful in other electoral settings, and to address other substantive questions. Consider, for example, studies using an RD approach to estimate partisan impacts on policy outcomes or roll call voting behavior—for example, Lee, Moretti, and Butler (2004), Leigh (2008), and Pettersson-Lidbom (2008). If the size of a party's vote reflects how much of a "mandate" it has from the voters, or if there is less than full party discipline in a legislature, then party control might have much larger impact on policy when the size of the majority is large than when the majority is razor thin. Similarly, the voting behavior of a legislator might depend on the size of his or her majority—for example, legislators who won with relatively comfortable majorities might feel more freedom to "vote their conscience" or vote to please their party's leadership, rather than "vote the district" on some issues. The technique employed in this article could be an important tool in assessing these and other hypotheses.

ACKNOWLEDGMENTS

We thank all researchers who responded to our survey.

REFERENCES

Angrist, Joshua D., and Miikka Rokkanen. Forthcoming. "Wanna Get Away? RD Identification Away from the Cutoff." *Journal of the American Statistical Association*.

Ansolabehere, Stephen, and James M. Snyder Jr. 2002. "The Incumbency Advantage in US Elections: An Analysis of State and Federal Offices, 1942–2000." *Election Law Journal* 1 (3): 315–38.

Ashworth, Scott, and Ethan Bueno de Mesquita. 2006. "Delivering the Goods: Legislative Particularism in Different Electoral and Institutional Settings." *Journal of Politics* 68 (1): 168–79.

Banks, Jeffrey S., and D. Roderick Kiewiet. 1989. "Explaining Patterns of Candidate Competition in Congressional Elections." *American Journal of Political Science* 33 (4): 997–1015.

Boas, Taylor C, and F. Daniel Hidalgo. 2011. "Controlling the Airwaves: Incumbency Advantage and Community Radio in Brazil." *American Journal of Political Science* 55 (4): 869–85.

Broockman, David E. 2009. "Do Congressional Candidates Have Reverse Coattails? Evidence from a Regression Discontinuity Design." *Political Analysis* 17 (4): 418–34.

Butler, Daniel Mark. 2009. "A Regression Discontinuity Design Analysis of the Incumbency Advantage and Tenure in the U.S. House." *Electoral Studies* 28 (1): 123–28.

Canon, David T. 1993. "Sacrificial Lambs or Strategic Politicians? Political Amateurs in US House Elections." *American Journal of Political Science* 37 (4): 1119–41.

Caughey, Devin M., and Jasjeet S. Sekhon. 2011. "Elections and the Regression Discontinuity Design: Lessons from Close US House Races, 1942–2008." *Political Analysis* 19 (4): 385–408.

Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein. 2010. "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design." *Quarterly Journal of Economics* 125 (1): 215–61.

Cox, Gary W., and Jonathan N. Katz. 1996. "Why Did the Incumbency Advantage in US House Elections Grow?" *American Journal of Political Science* 40 (2): 478–97.

Dal Bó, Ernesto, Pedro Dal Bó, and Jason Snyder. 2009. "Political Dynasties." *Review of Economic Studies* 76 (1): 115–42.

Desposato, Scott W., and John R. Petrocik. 2003. "The Variable Incumbency Advantage: New Voters, Redistricting, and the Personal Vote." *American Journal of Political Science* 47(1): 18–32.

Diamond, Alex, and Jasjeet Sekhon. 2013. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *Review of Economics and Statistics* 95 (3): 932–45.

DiNardo, John, and David S Lee. 2004. "Economic Impacts of New Unionization on Private Sector Employers: 1984–2001." *Quarterly Journal of Economics* 119 (4): 1383–1441.

Eggers, Andrew, Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder Jr. 2015. "On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: Evidence from Over 40,000 Close Races." *American Journal of Political Science* 59 (1): 259–74.

Eggers, Andrew C., and Jens Hainmueller. 2009. "MPs for Sale? Returns to Office in Postwar British Politics." *American Political Science Review* 103 (4): 513.

Erikson, Robert S. 1971. "The Advantage of Incumbency in Congressional Elections." *Polity* 3 (3): 395–405.

Ferreira, Fernando, and Joseph Gyourko. 2009. "Do Political Parties Matter? Evidence from U.S. Cities." *Quarterly Journal of Economics* 124 (1): 399–422.

Folke, Olle, and James M. Snyder Jr. 2012. "Gubernatorial Midterm Slumps." *American Journal of Political Science* 56 (4): 931–48.

Fouirnaies, Alexander, and Andrew B. Hall. 2014. "The Financial Incumbency Advantage: Causes and Consequences." *Journal of Politics* 76 (3): 711–24.

Gagliarducci, Stefano, and M. Daniele Paserman. 2012. "Gender Interactions within Hierarchies: Evidence from the Political Arena." *Review of Economic Studies* 79 (3): 1021–52.

Gerber, Elisabeth R, and Daniel J Hopkins. 2011. "When Mayors Matter: Estimating the Impact of Mayoral Partisanship on City Policy." *American Journal of Political Science* 55 (2): 326–39.

Grimmer, Justin, Eitan Hersh, Brian Feinstein, and Daniel Carpenter. 2012. "Are Close Elections Random?" Working Paper, Stanford University.

Hainmueller, Jens. 2012. "Entropy Balancing: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20 (1): 25–46.

Hainmueller, Jens, and Holger Lutz Kern. 2008. "Incumbency as a Source of Spillover Effects in Mixed Electoral Systems: Evidence from a Regression-Discontinuity Design." *Electoral Studies* 27 (2): 213–27.

Hirano, Shigeo, and James M. Snyder Jr. 2009. "Using Multimember District Elections to Estimate the Sources of the Incumbency Advantage." *American Journal of Political Science* 53 (2): 292–306.

Imbens, Guido W. 2010. "Better LATE than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)." *Journal of Economic Literature* 48 (2): 399–423.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.

Jacobson, G. C. 1989. "Strategic Politicians and the Dynamics of US House elections, 1946–86." *American Political Science Review* 83 (3): 773–93.

Jacobson, Gary C, and Samuel Kernell. 1983. *Strategy and Choice in Congressional Elections*. New Haven, CT: Yale University Press.

Kazee, Thomas A. 1994. "Ambition and Candidacy: Running as a Strategic Calculation." In Thomas A. Kazee, ed. *Who Runs for Congress? Ambition, Context, and Candidate Emergence*. Thousand Oaks, CA: CQ Press, 165–84.

Keele, Luke. 2010. "An Overview of R bounds: An R Package for Rosenbaum Bounds Sensitivity Analysis with Matched Data." White Paper. Columbus, OH.

King, Anthony S. 1997. *Running Scared: Why America's Politicians Campaign Too Much and Govern Too Little*. New York: Free Press.

Lee, David S. 2008. "Randomized Experiments from Non-Random Selection in US House Elections." *Journal of Econometrics* 142 (2): 675–97.

Lee, David S, Enrico Moretti, and Matthew J. Butler. 2004. "Do Voters Affect or Elect Policies? Evidence from the U.S. House." *Quarterly Journal of Economics* 119 (3): 807–59.

Leigh, Andrew. 2008. "Estimating the Impact of Gubernatorial Partisanship on Policy Settings and Economic Outcomes: A Regression Discontinuity Approach." *European Journal of Political Economy* 24 (1): 256–68.

Maisel, L. Sandy, and Walter J. Stone. 1997. "Determinants of Candidate Emergence in US House Elections: An Exploratory Study." *Legislative Studies Quarterly* 22 (1): 79–96.

Mann, Thomas. 1977. *Unsafe at Any Margin: Interpreting Congressional Elections*. Washington, DC: American Enterprise Institute.

Oppenheimer, Bruce I. 2005. "Deep Red and Blue Congressional Districts: The Causes and Consequences of Declining Party Competitiveness." In *Congress Reconsidered*, Vol. 8. Thousand Oaks, CA: CQ Press, 135–57.

Pettersson-Lidbom, Per. 2008. "Do Parties Matter for Economic Outcomes? A Regression-Discontinuity Approach." *Journal of the European Economic Association* 6 (5): 1037–56.

Rosenbaum, Paul R. 2002. *Observational Studies*. 2nd ed. New York: Springer.

Rosenbaum, Paul R. 2005. "Observational Study." In Brian Everitt and David Howell, eds. *Encyclopedia of Statistics in Behavioral Science*. Chichester: Wiley, 1809–14.

Snyder, Jason. 2005. "Detecting Manipulation in US House Elections." Unpublished manuscript.

Stein, Robert M., and Kenneth N. Bickers. 1997. *Perpetuating the Pork Barrel: Policy Subsystems and American Democracy*. Cambridge: Cambridge University Press.

Trounstine, Jessica. 2011. "Evidence of a Local Incumbency Advantage." *Legislative Studies Quarterly* 36 (2): 255–80.

Uppal, Yogesh. 2009. "The Disadvantaged Incumbents: Estimating Incumbency Effects in Indian State Legislatures." *Public Choice* 138 (1–2): 9–27.

Uppal, Yogesh. 2010. "Estimating Incumbency Effects in U.S. State Legislatures: A Quasi-Experimental Study." *Economics and Politics* 22 (2): 180–99.

Zaller, John. 1998. "Politicians as Prize Fighters: Electoral Selection and Incumbency Advantage." In John G. Geer, ed., *Politicians and Party Politics*. Baltimore: Johns Hopkins University Press, 125–85.