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

February 10, 2014

Abstract

The electoral regression discontinuity (RD) design is popular because it provides an unbiased, design-based estimate of the incumbency advantage with few assumptions. However, as is well known, the RD estimate is "local": it only identifies the effect in hypothetical districts with an exactly 50–50 tie between the Democratic and Republican candidates, and does not speak to the size of the incumbency advantage away from this threshold. There is significant uncertainty over the effect of incumbency in districts away from this threshold. Indeed, in a survey of political scientists that we administered, roughly equal numbers of respondents predict the effect to be either larger, smaller, or the same in less competitive districts. In this paper, we follow the method proposed in Angrist and Rokkanen (2013), employing a validated Conditional Independence Assumption that, unlike in typical cases, generates directly testable implications in the context of the RD. This technique allows us to estimate the average effect of incumbency in districts in a window around the threshold as large as 15 percentage points—i.e., 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 districts.

*Associate Professor. Department of Political Science and Graduate School of Business. jhain@stanford.edu, http://www.stanford.edu/~jhain/.

†Ph.D. Candidate. hall@fas.harvard.edu, andrewbenjaminhall.com.

‡Leroy B. Williams Professor of History and Political Science at Harvard University. jsnyder@gov.harvard.edu, http://scholar.harvard.edu/jsnyder/home.

1 Introduction

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 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 U.S. 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 U.S. elections. We discuss both of these in more detail below.

In this paper, we employ a new technique—developed in Angrist and Rokkanen (2013)—that allows us 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

[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.

[2] Angrist and Rokkanen (2013) apply the technique to an RD involving test scores and admission to exam schools in Boston.

explicitly testable implications. We apply the method to elections for U.S. 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.

The paper 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 U.S. statewide offices, presenting the results of the validity tests and the estimates of the incumbency advantage away from the threshold. Finally, we conclude.

2 Motivation

2.1 The RD Estimate Only Measures Incumbency Advantages In 50–50 Districts

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

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

[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 U.S. House, see Grimmer et al. (2012), Caughey and Sekhon (2011) and

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 U.S. elections are not extremely close (see Table 2 below), 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.

2.2 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 to utilize the direct office holder benefits since they are more vulnerable (e.g., Stein and Bickers 1997; Oppenheimer 2005; Ashworth and Bueno de Mesquita 2006).[5] 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. 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

Snyder (2005), but also see Eggers et al. (2013) for evidence that the assumption is widely plausible and that the U.S. House may be an exception and not evidence of sorting.

[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]Also, 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., Jacobson and Kernell (1983), Banks and Kiewiet (1989), Canon (1993), Kazee (1994), Cox and Katz (1996), and Maisel and Stone (1997) for arguments and evidence regarding strategic candidate entry and scare-off. Canon (1993: 1134-1135) 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."

expected vote for an incumbent by 5 percentage points. Then the scare-off effect might be much smaller in a 50–50 district 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 towards 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—i.e., they are all "running scared" (e.g., Mann 1977; King 1997). 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 U.S. House members and find evidence that the incumbency advantage is larger in areas that are "unfriendly" to the incumbent in terms of partisan affiliation, i.e., in areas where more voters are registered with the opposition party.

2.3 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 U.S. and first informed respondents that the party incumbency effect at the 50–50 threshold (based on RD estimates) is about a 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

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

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 email lists such as PolMeth (the email 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 incumbency effect is smaller, and about 33% of respondents think that the effect is about the same as in 50–50 districts. Among those who think that the incumbency advantage is larger, the average expected effect magnitude is 11.3 percentage points, about 32 percent 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 percent 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.

2.4 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.

[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

Do you expect the incumbency effect in these districts to be smaller or larger than at the 50% threshold?			
	Larger	Smaller	Same
% 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%	

Results from survey of 165 respondents recruited via Political Science email lists.

In Table 1 below we present a few summary statistics for different window sizes, for statewide elections in the U.S. over our period of study. The top panel (first 3 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 of the U.S. 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 1 shows the mean values of several variables—variables that are of interest in studies of the incumbency advantage—for the races that fall inside the various windows. Comparing the numbers in the first 7 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.

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	0.26	0.51	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 |Democ 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

Last row in top panel gives percent of state-years with at least one race in the given window.

3 Method

In this section we explain the method from Angrist and Rokkanen (2013) in the context of elections, and we describe the estimation strategies we use to implement the method using election data on U.S. statewide offices.

Let V_{ist} be the forcing variable, the Democratic vote share winning margin, in election i in state s in year t. The outcome, $Y_{is,t+1}$, is the Democratic vote share in the election at $t+1$. The treatment variable of interest is $D_{ist} \equiv 1\{V_{ist} > 0\}$, an indicator of Democratic victory at time t. As Angrist and Rokkanen (2013) point out, this means that D_{ist} is a deterministic function of the forcing variable, and therefore, in comparing treated and control units, V_{ist} 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 $V_{ist} > 0$ are treated with a Democratic incumbent and all elections with $V_{ist} < 0$ are treated with a Republican incumbent.

To get away from the discontinuity threshold, Angrist and Rokkanen (2013) propose gathering a set of control variables, X_{ist}, and imposing a conditional independence assumption (CIA), which asserts that

$$E[Y_{is,t+1}(D_{ist})|V_{its}, X_{ist}] = E[Y_{is,t+1}(D_{ist})|X_{ist}]$$

for $D_{ist} \in \{0, 1\}$. This says that once we condition on the set of covariates in X_{ist}, the potential outcomes are mean-independent of the forcing variable V_{ist}. In other words, by controlling for the set of covariates we break the correlation between the forcing variable and the potential outcomes and this ensures 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 $(Y_{is,t+1}(1), Y_{is,t+1}(0)) \perp\!\!\!\perp V_{ist}|X_{ist}$ and, given common support, the average treatment effect on the treated[11] (ATT) is therefore identified by a covariate adjusted comparison of the observed outcomes for treated and control units, as in

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

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 valid.

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

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

[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 Dem-winner units to Dem-loser units. Or, 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.

which means that we should see that

$$E[Y_{is,t+1}|V_{ist}, X_{ist}, D_{ist} = 1] = E[Y_{is,t+1}|X_{ist}, D_{ist} = 1] \quad (2)$$

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

$$E[Y_{is,t+1}|V_{ist}, X_{ist}, D_{ist} = 0] = E[Y_{is,t+1}|X_{ist}, D_{ist} = 0]. \quad (3)$$

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

$$Y_{is,t+1} = \beta_0 + \beta_1 V_{ist} + X_{ist} + \epsilon_{ist} \quad (4)$$

on each side of the threshold (in separate regressions) and testing for $\beta_1 = 0$. As Angrist and Rokkanen (2013) point out, it is likely that the CIA only holds within some window around the threshold.[12]

The intuition is as follows. The RD tells us that V_{ist} is the only omitted variable. If our control set successfully addresses the omitted variable bias, then when it and V_{ist} are included in the same regression on each side of the discontinuity, there should be no remaining correlation between V_{ist} and the outcome variable. If we find that there is no remaining correlation—i.e., 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.

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 $Y_{is,t+1}$ is regressed on the treatment indicator D_{ist} and the control variables included in the conditioning set X_{ist}. Second, to relax the linearity assumption we use one-to-one genetic matching with bias adjustment where we match each treated unit to the

[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.

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_{ist} (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).

4 Application to U.S. Statewide Offices

4.1 Data

We apply the method to the data on U.S. 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 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.6 in the Appendix.

4.2 Testing the Conditional Independence Assumption

Table 2 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, i.e., 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

[13] The method uses a genetic search algorithm to determine the weight that is given to each control variable in X_{ist} 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.

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 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. This indicates that the conditional independence assumption is plausible for these windows and conditioning sets. 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 dataset, 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.

Table 3 – Conditional Independence Tests. Presents CIA tests from equation 4 to the left of the dicontinuity (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.

	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	D=0	D=1	D=0	D=1	D=0	D=1
5	-0.08	0.32	-0.07	0.30	-0.02	0.23
	(0.33)	(0.31)	(0.33)	(0.31)	(0.31)	(0.28)
	N=439	*N=450*	*N=439*	*N=450*	*N=469*	*N=478*
10	0.05	0.03	0.05	0.03	0.09	0.06
	(0.11)	(0.12)	(0.11)	(0.12)	(0.11)	(0.12)
	N=834	*N=817*	*N=834*	*N=817*	*N=896*	*N=871*
15	0.27	0.00	0.27	0.00	0.28	0.06
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
	N=1167	*N=1137*	*N=1167*	*N=1137*	*N=1252*	*N=1205*
20	0.31	0.10	0.31	0.10	0.32	0.13
	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
	N=1386	*N=1398*	*N=1386*	*N=1398*	*N=1482*	*N=1481*
25	0.31	0.15	0.31	0.15	0.31	0.20
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
	N=1550	*N=1629*	*N=1550*	*N=1629*	*N=1652*	*N=1722*
30	0.29	0.18	0.29	0.18	0.29	0.22
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
	N=1652	*N=1800*	*N=1652*	*N=1800*	*N=1758*	*N=1895*
35	0.29	0.20	0.29	0.20	0.30	0.23
	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
	N=1733	*N=1916*	*N=1733*	*N=1916*	*N=1841*	*N=2020*
40	0.29	0.16	0.29	0.16	0.29	0.20
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	N=1780	*N=1998*	*N=1780*	*N=1998*	*N=1891*	*N=2110*

Robust standard errors in parentheses. V_{ist} and $Y_{is,t+1}$ measured in percentage points, 0–100.

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 above. 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 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.[15] 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.

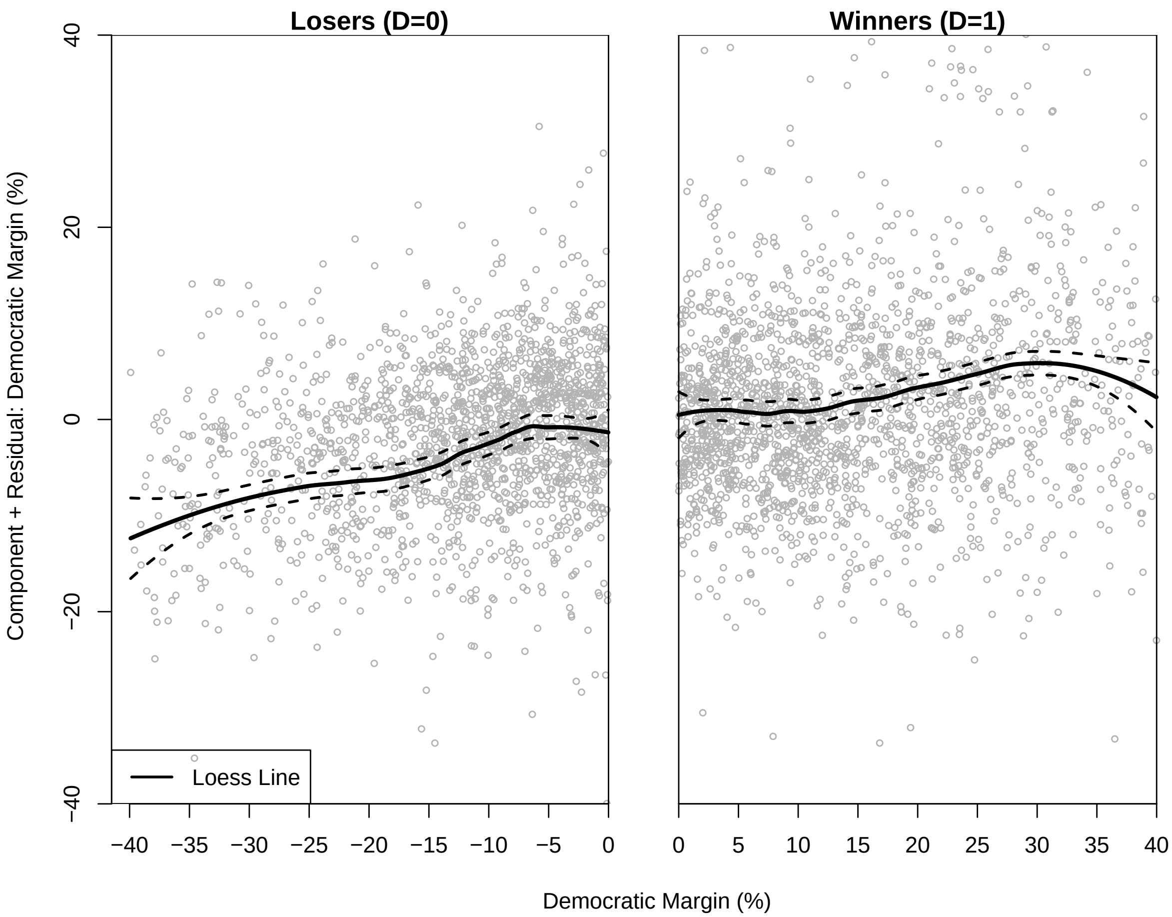
4.3 Incumbency Effects in Less Competitive Districts

The previous results suggest that the conditional independence assumption is valid for the 5% and 10% window and partially so for the 15% window. Therefore the incumbency effect for the districts 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 regression discontinuity design (RD) 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 districts, 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 robust across

[15] 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.

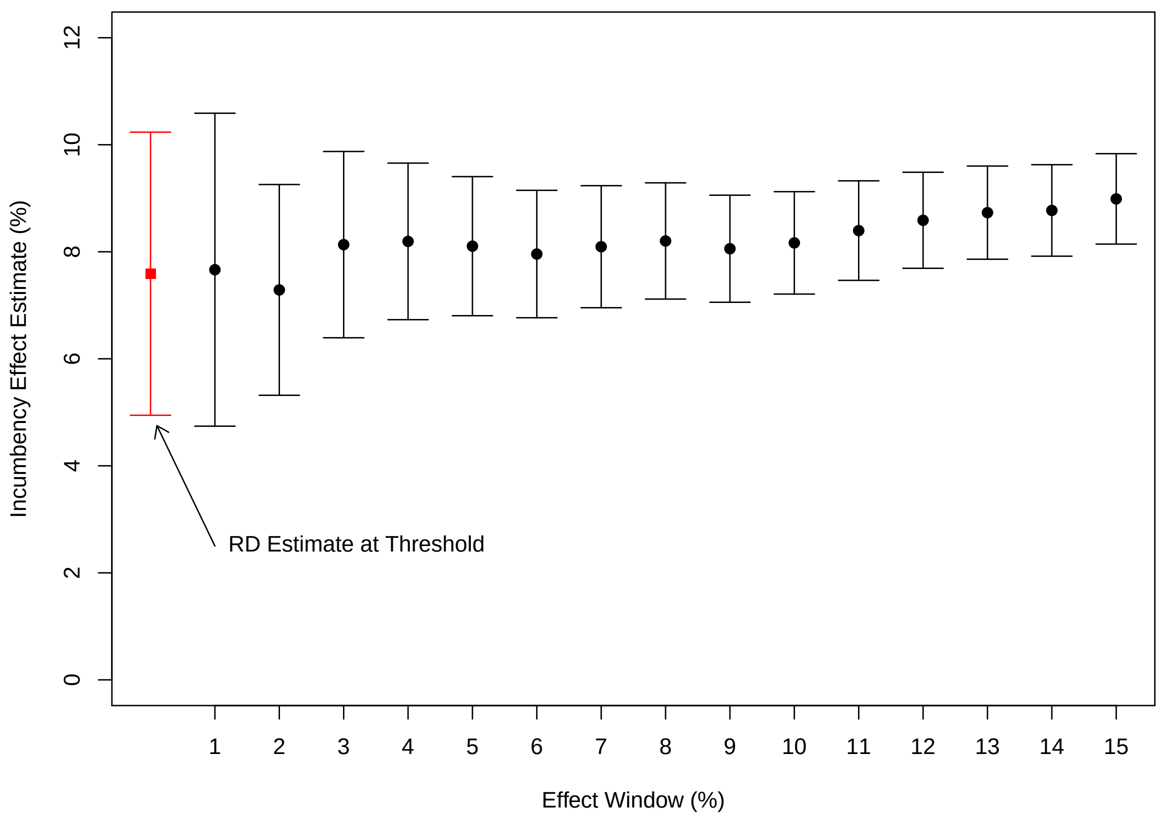
Figure 1 – Component Residual Plot to Test the Conditional Independence Assumption. Figure shows component residual plots that summarize the relationship between the forcing variable (Democratic winning margin) and the outcome (Democratic Vote Share in the next election) after partialling out the linear 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.



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.

In Figure 2 we also plot the incumbency effect estimates (with 95 percent 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.

Figure 2 – Incumbency Effects in Less Competitive Districts and at the Threshold. 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).



Looking at the magnitude of the effect—8.17% using OLS with control set 1 in a 10% window, for example—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 1).

4.4 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 includes hyper-competitive districts where the Democrats won with a razor thin margin of just over zero percentage points, as well as relatively safer districts where the Democrat 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 restricts the sample of winners to districts where the Democrats 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 Democrats lost with margins between 0-10%.

Table 5 presents the results. We find that the effect estimates are very similar to the previous estimates. Even when we exclude districts where the Democrats barely won, the effect magnitudes are still about 8-9 percentage points. These results are robust across all three windows, covariate adjustments, and conditioning sets.[16] These findings suggest that the incumbency effect in less competitive districts 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.

[16] As in the previous analysis, balance tests for these control strategies are available in the Appendix.

Table 4 – Incumbency Effects in Less Competitive Districts and at the Threshold. The top panel presents incumbency effect estimates in less competetive districts based on the conditional independence assumption for different windows and covariate adjustment methods. The bottom panel presents for comparison the incumbency effect estimates at the threshold based on a regression discontinuity design for different bandwidths.

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	8.10	7.77	8.42	8.10	7.69	8.42	7.67	7.68	7.99
	(0.65)	(0.73)	(0.67)	(0.65)	(0.76)	(0.67)	(0.62)	(0.71)	(0.64)
	N=889	*N=889*	*N=889*	*N=889*	*N=889*	*N=889*	*N=947*	*N=947*	*N=947*
10	8.17	7.53	8.41	8.17	8.19	8.41	7.97	8.40	8.23
	(0.48)	(0.65)	(0.53)	(0.48)	(0.62)	(0.53)	(0.46)	(0.69)	(0.49)
	N=1651	*N=1651*	*N=1651*	*N=1651*	*N=1651*	*N=1651*	*N=1767*	*N=1767*	*N=1767*
15	8.99	9.14	9.29	8.99	8.99	9.29	8.95	9.85	9.27
	(0.42)	(0.63)	(0.49)	(0.42)	(0.61)	(0.49)	(0.40)	(0.57)	(0.43)
	N=2304	*N=2304*	*N=2304*	*N=2304*	*N=2304*	*N=2304*	*N=2457*	*N=2457*	*N=2457*

Incumbency Effect at the Threshold (RD estimates)

Bandwidth	Local Linear	Local Linear	Local Linear
1	9.22	9.22	8.66
	(3.60)	(3.60)	(3.36)
	N=179	*N=179*	*N=192*
2	8.26	8.26	7.98
	(2.28)	(2.28)	(2.13)
	N=362	*N=362*	*N=385*
5	7.59	7.59	7.28
	(1.32)	(1.32)	(1.24)
	N=889	*N=889*	*N=947*

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. $Y_{is,t+1}$ measured in percentage points, 0–100.

Table 5 – Incumbency Effects in Less Competitive Districts Further Away from the Threshold. Presents incumbency effect estimates in less competetive 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.

	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	8.12	7.96	8.25	8.12	7.74	8.26	8.08	9.04	8.26
	(0.64)	(0.85)	(0.70)	(0.64)	(0.86)	(0.70)	(0.61)	(0.84)	(0.65)
	N=1201	*N=1201*	*N=1201*	*N=1201*	*N=1201*	*N=1201*	*N=1289*	*N=1289*	*N=1289*
5-15	8.26	8.18	8.41	8.26	8.49	8.41	8.37	9.33	8.55
	(0.52)	(0.74)	(0.60)	(0.52)	(0.69)	(0.60)	(0.50)	(0.71)	(0.54)
	N=1521	*N=1521*	*N=1521*	*N=1521*	*N=1521*	*N=1521*	*N=1623*	*N=1623*	*N=1623*
10-15	8.50	8.32	8.59	8.50	8.28	8.59	8.76	8.55	8.90
	(0.67)	(1.02)	(0.76)	(0.67)	(0.88)	(0.76)	(0.65)	(0.85)	(0.69)
	N=1154	*N=1154*	*N=1154*	*N=1154*	*N=1154*	*N=1154*	*N=1230*	*N=1230*	*N=1230*

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. Margin 5-10 means that only districts where the Democrats won with margins bewteen 5 and 10% are included. For all margins, the sample of losers consists of districts where the Demoracts lost with margins between -10 and 0%. $Y_{is,t+1}$ measured in percentage points, 0–100.

5 Conclusion

In this paper, we have applied a new technique from the econometrics literature to investigate the causal effect of incumbency on U.S. statewide electoral outcomes away from the 50–50 threshold at which the RD estimates it. We find that the incumbency advantage continues to be just as large when measured in a window as wide as 15% around the discontinuity threshold. As we showed, estimating the effect in this 15% window includes observations from over 90% of all state-decades and more than half (51.7%) of all races, whereas the RD estimate includes less than 24% of state-decades and less than 1.9% 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 U.S. elections.

From a substantive standpoint, the results show that incumbents continue to enjoy the same large electoral advantage in somewhat less competitive districts. There does not appear to be any interaction between incumbency and district safety within the window of elections we study. Even when incumbents win by relatively wide margins—margins as big as 15 percentage points—they do not receive a larger electoral advantage in the next campaign. Incumbents thus do not appear to reduce effort after winning by relatively wide margins, as some have predicted. Instead, incumbents continue to enjoy the same advantage in all elections, on average, whether they win by 1 percentage point or 15.

Finally, the technique we have applied to generalize the RD estimate of the incumbency advantage is likely to be useful in many 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—e.g., 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 the 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 – e.g., 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.

References

Angrist, Joshua D., and Miikka Rokkanen. 2013. "Wanna Get Away? RD Identification Away From The Cutoff." NBER Working Paper.

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): 315–338.

Ashworth, Scott, and Ethan Bueno de Mesquita. 2006. "Delivering the Goods: Legislative Particularism in Different Electoral and Institutional Settings." *Journal of Politics* 68(1): 168–179.

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): 869–885.

Broockman, David E. 2009. "Do Congressional Candidates Have Reverse Coattails? Evidence From A Regression Discontinuity Design." *Political Analysis* 17(4): 418–434.

Butler, Daniel Mark. 2009. "A Regression Discontinuity Design Analysis of the Incumbency Advantage and Tenure in the U.S. House." *Electoral Studies* 28(1): 123–128.

Canon, David T. 1993. "Sacrificial Lambs or Strategic Politicians? Political Amateurs in US House Elections." *American Journal of Political Science* 37(4): 1119–1141.

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." *The Quarterly Journal of Economics* 125(1): 215–261.

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): 478–497.

Dal Bó, Ernesto, Pedro Dal Bó, and Jason Snyder. 2009. "Political Dynasties." *The Review of Economic Studies* 76(1): 115–142.

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): 932–945.

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

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

Eggers, Andrew, Olle Folke, Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder Jr. 2013. "On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: Evidence From Over 40,000 Close Races." Working Paper.

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." *The 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): 931–948.

Fouirnaies, Alexander, and Andrew B. Hall. N.d. "The Financial Incumbency Advantage: Causes and Consequences." *Journal of Politics*. Forthcoming.

Gagliarducci, Stefano, and M Daniele Paserman. 2012. "Gender Interactions Within Hierarchies: Evidence from the Political Arena." *The Review of Economic Studies* 79(3): 1021–1052.

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): 326–339.

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

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): 213–227.

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): 615–635.

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

Kazee, Thomas A. 1994. "Ambition and Candidacy: Running as a Strategic Calculation." In *Who Runs for Congress? Ambition, Context, and Candidate Emergence*, ed. Thomas A. Kazee. CQ Press pp. 165–184.

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

Lee, David S. 2008. "Randomized Experiments From Non-Random Selection in US House Elections." *Journal of Econometrics* 142(2): 675–697.

Lee, David S, Enrico Moretti, and Matthew J. Butler. 2004. “Do Voters Affect or Elect Policies? Evidence from the U.S. House.” *The Quarterly Journal of Economics* 119(3): 807–859.

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): 256–268.

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.* 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 CQ Press pp. 135–57.

Pettersson-Lidbom, Per. 2008. “Do Parties Matter For Economic Outcomes? A Regression-Discontinuity Approach.” *Journal of the European Economic Association* 6(5): 1037–1056.

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 University Press.

Trounstine, Jessica. 2011. “Evidence of a Local Incumbency Advantage.” *Legislative Studies Quarterly* 36(2): 255–280.

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 & Politics* 22(2): 180–199.

Zaller, John. 1998. “Politicians as Prize Fighers: Electoral Selection and Incumbency Advantage.” In *Politicians and Party Politics*, ed. John G. Geer. Johns Hopkins University Press pp. 125–185.

1 Appendix

1.1 Survey

The survey intro was:

- "Consider the party incumbency effect in elections for statewide offices in the U.S. in recent decades.

  When using a regression discontinuity design that exploits variation from close elections we find that party incumbency increases the two-party vote share in the next election by around 8-9 percentage points on average.

  As is well known, this effect estimate only refers to very close elections that are decided within a narrow window around the 50% vote share threshold of winning, e.g. elections in which the party barley won with a vote of 50.5%.

  Here we are interested in your expectation of what the party incumbency effect might be in districts where the winner received a vote share that was substantially higher than the 50% threshold."

The first question was:

- "Consider the party incumbency effect in districts where the winner received between 50% and 60% of the vote.

  Do you expect the party incumbency effect in these districts to be smaller or larger than in districts right at the 50% threshold?"

- Answer options:
  - incumbency effect is smaller than at the 50% threshold
  - incumbency effect is about the same as at the 50% threshold
  - incumbency effect is smaller than at the 50% threshold

The next question was:

- "What magnitude do you expect for the party incumbency effect in districts where the winner received between 50% and 60% of the vote?

  Please move the slider to your expected effect size (e.g. 1 means you expect a 1 percentage points increase in incumbent party vote share). As a reminder: the effect at the 50% threshold is around 8-9 percentage points."

1.2 Balance Checks for Table 3

Here we check the overlap in the covariate distributions. Tables A.1 and A.2 below summarize for the 5% and the 10% window the covariate balance in the raw and adjusted data for the conditioning set 1, the most extensive set of control variables. Overall we find that there is sufficient overlap in both of these windows for which we found the conditional independence assumption appears valid. While there are significant imbalances in the raw data, these imbalances largely disappear in the unmatched or reweighed data; the means are close together, the p-values from the difference in means tests are all above .05, and the variance ratios are close to one. Taken together these results suggests that there is enough covariate overlap in these windows to allow for a robust identification; a fact that is consistent with the finding that the incumbency effects estimates presented above do not vary much across the different adjustment methods.

Table A.1 – Balance Before and After Covariate Adjustment (Window 5%)

Adjustment	Covariate	Mean Tr	Mean Co	S.Diff	T.pval	Var.Ratio
Unmatched	Dem Share t-1	5.49	−1.12	0.45	0.00	1.26
Unmatched	Dem Share t-2	3.30	1.13	0.13	0.16	1.29
Unmatched	Normal Votet t-1	52.02	50.69	0.21	0.02	0.96
Unmatched	Normal Vote t-2	51.59	51.05	0.08	0.41	0.93
Unmatched	Midterm Slump t	−0.16	−0.18	0.03	0.72	1.03
After Genetic Matching	Dem Share t-1	5.49	4.88	0.03	0.06	1.36
After Genetic Matching	Dem Share t-2	3.30	3.46	−0.01	0.68	1.12
After Genetic Matching	Normal Votet t-1	52.02	52.42	−0.05	0.06	0.96
After Genetic Matching	Normal Vote t-2	51.59	51.47	0.01	0.46	1.05
After Genetic Matching	Midterm Slump t	−0.16	−0.18	0.02	0.07	1.00
After Entropy Balancing	Dem Share t-1	5.49	5.49	0.00	1.00	1.19
After Entropy Balancing	Dem Share t-2	3.30	3.30	0.00	1.00	1.15
After Entropy Balancing	Normal Votet t-1	52.02	52.02	0.00	1.00	0.82
After Entropy Balancing	Normal Vote t-2	51.59	51.59	0.00	1.00	0.87
After Entropy Balancing	Midterm Slump t	−0.16	−0.16	0.00	1.00	0.99

S.Diff=Standardized difference in means; T-pval=p-value from difference in means test; Var.Ratio: Ratio of variances

Table A.2 – Balance Before and After Covariate Adjustment (Window 10%)

Adjustment	Covariate	Mean Tr	Mean Co	S.Diff	T.pval	Var.Ratio
Unmatched	Dem Share t-1	6.56	−1.51	0.53	0.00	1.17
Unmatched	Dem Share t-2	5.28	−0.38	0.34	0.00	1.53
Unmatched	Normal Votet t-1	52.79	49.98	0.44	0.00	1.05
Unmatched	Normal Vote t-2	52.43	50.24	0.32	0.00	1.13
Unmatched	Midterm Slump t	−0.18	−0.11	−0.14	0.04	0.95
After Genetic Matching	Dem Share t-1	6.56	6.47	0.00	0.59	1.03
After Genetic Matching	Dem Share t-2	5.28	5.35	0.00	0.71	1.05
After Genetic Matching	Normal Votet t-1	52.79	52.82	0.00	0.82	1.01
After Genetic Matching	Normal Vote t-2	52.43	52.37	0.01	0.54	1.05
After Genetic Matching	Midterm Slump t	−0.18	−0.17	−0.01	0.41	1.01
After Entropy Balancing	Dem Share t-1	6.56	6.56	0.00	1.00	0.76
After Entropy Balancing	Dem Share t-2	5.28	5.28	0.00	1.00	1.03
After Entropy Balancing	Normal Votet t-1	52.79	52.79	0.00	1.00	0.71
After Entropy Balancing	Normal Vote t-2	52.43	52.43	0.00	1.00	0.84
After Entropy Balancing	Midterm Slump t	−0.18	−0.18	0.00	1.00	0.98

S.Diff=Standardized difference in means; T-pval=p-value from difference in means test; Var.Ratio: Ratio of variances

1.3 Balance Checks for Table 4

The Tables A.3-A.5 below report the balance statistics for the samples used for the effect estimation in Table 4, where we exclude the observations that are right above the threshold. Again, we find that there is sufficient overlap in both of these windows for which we found the conditional independence assumption appears valid.

Table A.3 – Balance Before and After Covariate Adjustment (Margins of 5-10%)

Adjustment	Covariate	Mean Tr	Mean Co	S.Diff	T.pval	Var.Ratio
Unmatched	Dem Share t-1	7.89	−1.51	0.63	0.00	1.20
Unmatched	Dem Share t-2	7.70	−0.38	0.50	0.00	1.63
Unmatched	Normal Votet t-1	53.75	49.98	0.58	0.00	1.15
Unmatched	Normal Vote t-2	53.46	50.24	0.46	0.00	1.28
Unmatched	Midterm Slump t	−0.20	−0.11	−0.18	0.04	0.94
After Genetic Matching	Dem Share t-1	7.89	7.85	0.00	0.81	1.03
After Genetic Matching	Dem Share t-2	7.70	7.43	0.01	0.59	1.19
After Genetic Matching	Normal Votet t-1	53.75	53.66	0.01	0.63	1.05
After Genetic Matching	Normal Vote t-2	53.46	53.36	0.01	0.63	1.09
After Genetic Matching	Midterm Slump t	−0.20	−0.19	−0.01	0.59	1.02
After Entropy Balancing	Dem Share t-1	7.89	7.89	0.00	1.00	0.72
After Entropy Balancing	Dem Share t-2	7.70	7.70	0.00	1.00	0.94
After Entropy Balancing	Normal Votet t-1	53.75	53.75	0.00	1.00	0.67
After Entropy Balancing	Normal Vote t-2	53.46	53.46	0.00	1.00	0.81
After Entropy Balancing	Midterm Slump t	−0.20	−0.20	0.00	1.00	0.97

S.Diff=Standardized difference in means; T-pval=p-value from difference in means test; Var.Ratio: Ratio of variances

Table A.4 – Balance Before and After Covariate Adjustment (Margins 5-15%)

Adjustment	Covariate	Mean Tr	Mean Co	S.Diff	T.pval	Var.Ratio
Unmatched	Dem Share t-1	8.48	−1.51	0.65	0.00	1.24
Unmatched	Dem Share t-2	9.14	−0.38	0.58	0.00	1.46
Unmatched	Normal Votet t-1	53.86	49.98	0.60	0.00	1.08
Unmatched	Normal Vote t-2	53.81	50.24	0.50	0.00	1.21
Unmatched	Midterm Slump t	−0.16	−0.11	−0.10	0.15	0.95
After Genetic Matching	Dem Share t-1	8.48	8.37	0.00	0.53	1.02
After Genetic Matching	Dem Share t-2	9.14	9.03	0.00	0.55	1.02
After Genetic Matching	Normal Votet t-1	53.86	53.78	0.01	0.60	1.02
After Genetic Matching	Normal Vote t-2	53.81	53.77	0.00	0.78	1.13
After Genetic Matching	Midterm Slump t	−0.16	−0.15	−0.01	0.50	1.02
After Entropy Balancing	Dem Share t-1	8.48	8.48	0.00	1.00	0.72
After Entropy Balancing	Dem Share t-2	9.14	9.14	0.00	1.00	0.77
After Entropy Balancing	Normal Votet t-1	53.86	53.86	0.00	1.00	0.62
After Entropy Balancing	Normal Vote t-2	53.81	53.81	0.00	1.00	0.75
After Entropy Balancing	Midterm Slump t	−0.16	−0.16	0.00	1.00	0.97

S.Diff=Standardized difference in means; T-pval=p-value from difference in means test; Var.Ratio: Ratio of variances

Table A.5 – Balance Before and After Covariate Adjustment (Margins 10-15%)

Adjustment	Covariate	Mean Tr	Mean Co	S.Diff	T.pval	Var.Ratio
Unmatched	Dem Share t-1	9.16	−1.51	0.70	0.00	1.27
Unmatched	Dem Share t-2	10.79	−0.38	0.72	0.00	1.26
Unmatched	Normal Votet t-1	53.98	49.98	0.63	0.00	1.02
Unmatched	Normal Vote t-2	54.21	50.24	0.57	0.00	1.14
Unmatched	Midterm Slump t	−0.12	−0.11	−0.02	0.83	0.96
After Genetic Matching	Dem Share t-1	9.16	9.14	0.00	0.91	1.01
After Genetic Matching	Dem Share t-2	10.79	10.75	0.00	0.90	1.02
After Genetic Matching	Normal Votet t-1	53.98	53.95	0.00	0.84	1.03
After Genetic Matching	Normal Vote t-2	54.21	54.24	0.00	0.94	1.20
After Genetic Matching	Midterm Slump t	−0.12	−0.11	−0.01	0.58	1.03
After Entropy Balancing	Dem Share t-1	9.16	9.16	0.00	1.00	0.71
After Entropy Balancing	Dem Share t-2	10.79	10.79	0.00	1.00	0.61
After Entropy Balancing	Normal Votet t-1	53.98	53.98	0.00	1.00	0.57
After Entropy Balancing	Normal Vote t-2	54.21	54.21	0.00	1.00	0.67
After Entropy Balancing	Midterm Slump t	−0.12	−0.12	0.00	1.00	0.97

S.Diff=Standardized difference in means; T-pval=p-value from difference in means test; Var.Ratio: Ratio of variances

1.4 Data

Here we provide more information about the statewide elections dataset we employ in the paper. Table A.6 shows the number of data points used in the analysis with Control Set 3, the most parsimonious of the control sets. Specifically, each cell is the total number of data points entering the sample for a particular state and office, across the full range of values of the RD bandwidth or CIA window. The table does *not* count data points that have missing values for the outcome variable or for any of the control variables, so as to correspond precisely to the regression results reported. Note that some states have 0's in some columns reflecting the fact that those states do not hold elections for those offices (e.g., Alaska does not elect its attorney general and New Jersey does not elect any state executive office other than governor).

Table A.6 – Observations in Data Set, by State and Office. Each cell provides the total number of data points in the dataset used for analysis, subset to observations with no missing values for Control Set 3.

State	# Att Genl	# Auditor	# Gov	# LT Gov	# Senate	# Sec State	# Treasurer	Min Year	Max Year
AK	0	0	9	0	15	0	0	1960	2006
AL	14	13	14	15	16	13	13	1950	2006
AR	24	24	25	25	17	25	25	1948	2006
AZ	20	10	21	0	19	21	20	1948	2006
CA	15	0	14	15	19	15	14	1950	2006
CO	18	6	18	7	19	18	18	1948	2006
CT	15	0	15	8	19	16	16	1948	2006
DE	15	24	15	15	21	0	24	1948	2008
FL	15	0	15	0	20	12	13	1950	2006
GA	15	0	13	15	18	15	5	1950	2006
HI	0	0	10	0	15	0	0	1962	2006
IA	22	21	22	16	19	22	22	1948	2006
ID	15	14	15	15	18	14	12	1950	2006
IL	15	4	14	4	19	14	18	1948	2006
IN	15	20	15	5	19	20	20	1948	2008
KS	22	11	22	12	18	22	23	1948	2006
KY	15	15	15	10	19	15	15	1950	2007
LA	5	5	5	5	7	5	5	1950	1968
MA	20	18	20	10	19	19	19	1948	2006
MD	15	0	15	0	20	0	0	1950	2006
ME	0	0	16	0	20	0	0	1948	2006
MI	20	7	20	7	20	20	7	1948	2006
MN	19	15	18	9	20	18	16	1948	2006
MO	15	15	15	15	20	15	15	1950	2008
MS	15	14	13	13	18	14	15	1948	2007
MT	15	14	15	5	20	14	5	1948	2008
NC	15	15	15	15	19	15	13	1950	2008
ND	20	20	20	10	19	20	20	1948	2008
NE	19	20	20	10	20	20	18	1948	2006
NH	0	0	32	0	19	0	0	1948	2010
NJ	0	0	15	0	20	0	0	1948	2006
NM	21	20	21	7	20	21	20	1948	2006
NV	15	0	15	15	20	14	14	1950	2006
NY	15	0	15	0	19	0	0	1950	2006
OH	18	15	18	9	20	18	18	1948	2006
OK	15	15	15	15	19	6	15	1950	2006
OR	12	0	16	0	19	16	16	1948	2010
PA	7	15	15	4	20	4	14	1950	2008
RI	27	0	27	27	21	27	27	1948	2006
SC	14	0	15	15	18	14	12	1950	2006
SD	22	22	22	12	19	22	22	1948	2006
TN	0	0	13	0	20	0	0	1948	2006
TX	20	0	22	20	20	0	14	1948	2006
UT	16	15	14	0	20	6	14	1950	2008
VA	13	0	13	14	12	0	0	1948	2006
VT	24	26	31	32	19	28	21	1948	2010
WA	14	15	15	15	20	15	15	1950	2008
WI	20	0	21	10	20	21	21	1948	2006
WV	15	15	15	0	20	14	15	1948	2008
WY	0	15	15	0	21	13	15	1948	2008