Online Appendix
Disentangling the Personal and Partisan Incumbency Advantages: Evidence from Close Elections and Term Limits

Anthony Fowler and Andrew B. Hall

Contents

1 More Details on the Estimation Procedure **2**
- 1.1 RD Estimation 2
- 1.2 Estimation of Run Probabilities 4

2 Robustness of the Personal and Partisan Estimates **4**

3 Are Expiring Cases Meaningfully Different from Non-expiring Cases? **8**

4 Are Expiring and Non-expiring Elections Comparable? **11**

5 Replicating Lee (2008) Estimates in Our Sample and Testing for Sorting **11**

6 Comparison to Other Estimates of the Incumbency Advantage **16**

7 Discussion of Competitive Elections and our Local Average Estimates **17**

8 Considering Bias from Replacement Candidate Quality **19**
- 8.1 Model of Candidate Quality and Electoral Selection 21
- 8.2 Perturbing the Model to Obtain Non-Zero Average Quality at the RD Threshold 23
- 8.3 Equal Candidate Quality with Unequal Variances 24
- 8.4 Incumbent Quality in Close Elections 25

1 More Details on the Estimation Procedure

As mentioned in the paper, our empirical strategy only exploits within-chamber variation between expiring and non-expiring elections. As a result, the fact that some states or legislative chambers may be systematically different from others does not pose a problem for our inferences. In this section, we provide more detail on the specific estimation of our parameters. We estimate the RD effect in expiring and non-expiring cases, respectively, with a single regression model that includes state-chamber fixed effects. Similarly, we estimate the probability that the winner of a close election runs again in both cases with a single regression that also includes state-chamber fixed effects. In the following sub-sections, we detail the estimation and show that our results are robust to numerous different specifications.

1.1 RD Estimation

In Table A1, we present our estimates of the RD effect in both expiring and non-expiring cases. For each dependent variable, Democratic Vote Share and Democratic Victory at time $t+1$, we regress the outcome on Democratic Victory at time t (our dummy treatment variable), a dummy variable indicating expiring cases, and an interaction between the two. We also include a fourth order polynomial of Democratic Vote Share at time t (the running variable) along with interactions of these variables with the expiring variable. This allows the outcomes to vary across the running variable differently for both the expiring and non-expiring cases. The running variable is coded to range from -.5 to .5, so that the threshold is at zero and so that the constant term can be interpreted as the expected value of the dependent variable for a non-expiring case where the Democratic Party barely-lost the previous election.

The coefficient on the variable Dem Victory (t) represents the RD effect for non-expiring cases, the coefficient on the interactive variable Dem Victory (t) * Expiring , represents the difference between the RD effect in expiring and non-expiring cases, and the sum of those two coefficients represents the RD effect for expiring cases. Before running the regression, the data was de-meaned by state-chamber, meaning that for each observation and for each variable we subtracted the state-chamber mean of that variable and added back the population mean (this is automated with the areg command in Stata). For this reason, the difference in the RD estimate between expiring and non-expiring cases is driven only by variation within state-chambers. As a result, differences between chambers and legislative settings are not a threat to internal validity.

Table A1. Estimating The RD Effects In Expiring And Nonexpiring Elections

	(1) DV = Dem Vote share (t+1)	(2) DV = Dem Victory (t+1)
Dem Victory (t)	**.078**	**.480**
	(.011)	(.025)
Expiring	.007	.113
	(.015)	(.042)
Dem Victory (t) * Expiring	**−.054**	**−.328**
	(.025)	(.075)
Dem Vote share (t)	1.021	1.209
	(.065)	(.108)
Dem Vote share (t)^2	−.078	−.042
	(.244)	(.362)
Dem Vote share (t)^3	−3.537	−7.115
	(.778)	(.925)
Dem Vote share (t)^4	3.958	6.418
	(2.334)	(2.824)
Dem Vote share (t)* Expiring	−.417	.444
	(.115)	(.273)
Dem Vote share (t)^2 * Expiring	.393	1.258
	(.424)	(.801)
Dem Vote share (t)^3 * Expiring	5.059	2.651
	(.955)	(1.477)
Dem Vote share (t)^4 * Expiring	−6.048	−8.378
	(3.129)	(4.485)
Constant	.457	.216
	(.006)	(.013)
Data de-meaned by state-chamber	X	X
Observations	5576	5576
R-squared	.48	.55
SER	.17	.33

Robust standard errors in parentheses.
*The table presents the full models used to estimate the RD quantities. The first column employs vote share as the dependent variable and the second column employs victory as the dependent variable. The coefficients on Dem Victory show the RD effects for non-expiring elections. The coefficients on Dem Victory*Expiring show the extent to which the RD effects change as we move from non-expiring to expiring elections. Therefore, if we sum these two coefficients, we obtain the RD effect for expiring elections. The models include state-chamber fixed effects (data de-meaned by state-chamber), so only the within-chamber differences between expiring and non-expiring cases will contribute to the estimates.*

1.2 Estimation of Run Probabilities

Table A2 presents our estimates of the probability that the winner of a close election runs again in the next election for expiring and non-expiring cases. In order to focus on variation within state-chambers and to avoid concerns about different electoral settings, we estimate both probabilities in a single regression model, after de-meaning the data by state-chamber (again this is done with the areg command in Stata). The dependent variable is a dummy variable indicating whether the winner sought reelection. The independent variables include a fourth-order polynomial of the winner's margin of victory (coded from 0 to .5), interactions of those variables with the expiring indicator variable, and the expiring variable itself. We want to know the limit of the probability that the winner runs again as the winner's margin of victory approaches zero, so the constant coefficient in the regression represents that probability for the non-expiring cases. The coefficient on expiring represents the difference in this probability between the expiring and non-expiring cases, and we can calculate this limit for the expiring cases by adding those two coefficients.

2 Robustness of the Personal and Partisan Estimates

As discussed in the paper, several recent studies challenge the use of RD in electoral settings (Caughey and Sekhon 2012; Grimmer et al. 2012). For example, we might worry that incumbents are better at winning close election or there may be some strategic sorting around very close elections. Table A3 shows that our results are robust to these challenges. The first row presents our estimates of the personal and partisan incumbency advantages with respect to both vote share and probability of victory, as shown in Table 2. The next row shows how these results change if we include covariates in our estimation. We control for both the vote share and electoral outcome in the previous election and see that our estimates of the personal and partisan advantages are virtually unchanged. Then, in the final row of Table A3, we conduct a donut RD analysis, where we exclude all elections from our sample where the margin of victory was less than half a percentage point. If there is strategic sorting around the threshold, then it will occur for these extremely close elections. By removing these observations, we can still generate unbiased estimates without the concern that strategic sorting plagues the results. Again, the results are virtually unchanged, suggesting that strategic sorting is not a concern.

For all of our previous analyses, we include a fourth order polynomial of the running variable. We could similarly estimate the personal and partisan incumbency advantages using different specifications. In Table A4, we present estimates of the personal and partisan advantages using 1st, 2nd, 3rd, and 4th order polynomials. We also vary whether we estimate

Table A2. Estimating Run Probabilities In Expiring And Nonexpiring Elections

	DV = Winner Runs Again (t+1)
Winner Margin	.273
	(.880)
Winner Margin^2	−6.987
	(9.081)
Winner Margin^3	18.963
	(33.638)
Winner Margin^4	−14.036
	(39.762)
Winner Margin * Expiring	−6.808
	(1.764)
Winner Margin^2 * Expiring	48.315
	(15.568)
Winner Margin^3 * Expiring	−129.596
	(52.123)
Winner Margin^4 * Expiring	118.789
	(57.472)
Expiring	**−.308**
	(.065)
Constant	**.678**
	(.025)
Data de-meaned by state-chamber	X
Observations	5576
R-squared	.24
SER	.44

Robust standard errors in parentheses.
The model estimates the probability that a barely-winner seeks reelection in both expiring and non-expiring cases. The constant term indicates the probability that a barely-winner will seek reelection in the non-expiring cases. The coefficient on Expiring indicates the extent to which this probability changes as we move from non-expiring to expiring cases. Therefore, we can sum these two coefficients to obtain the probability that a barely-winner will seek reelection in the expiring cases. As before, the model includes state-chamber fixed effects (data de-meaned by state-chamber), so only the within-chamber differences between expiring and non-expiring cases will contribute to the estimates.

Table A3. Robustness Of The Personal And Partisan Estimates To Challenges Against RD

	Vote share		Victory	
	Personal	Partisan	Personal	Partisan
Estimate	.088	−.020	.533	−.121
w/ Controls	.088	−.016	.529	−.101
Donut RD	.077	−.013	.425	−.043

The table shows the robustness of our findings to alternative specifications and challenges of sorting in close elections. Row 1 shows our estimates of the personal and partisan advantages. Row 2 shows the same estimates if we include controls for previous vote share and previous electoral victory. Row 3 shows the same estimates if we exclude all elections where the margin of victory was less than 0.5 percentage points. In each case, the results are largely unchanged.

the same model for the entire range of data or if we estimate different models on each side of the threshold. For each of those 8 specifications, we choose the appropriate bandwidth that provides efficient estimates without sacrificing consistency. Table A4 shows that our results are robust to these alternative specifications. For each specification, the personal advantage is significantly larger than the partisan advantage, and the substantive size of the effects are similar across all models.

In the final row of the table, we employ a separate technique. Instead of using OLS with a function of the running variable at a chosen bandwidth, we use the automated technique from Calonico, Cattaneo, and Titiunik (2014), as implemented in the Stata package rdrobust. This technique selects a data-driven bandwidth and estimates effects using a kernel-weighted local linear function of the running variable. Applying this technique in our context has several complexities. As before, we need to estimate four parameters: the probability expiring and non-expiring winners run in the next election, and the RD estimate on either vote share or victory in expiring and non-expiring cases. First, we must residualize the outcome variables because rdrobust does not accept covariates as inputs. Thus the first step is to regress vote share or victory and the indicator for the winner running again on state-chamber dummies. Next, we estimate each of the four quantities separately using rdrobust with all default options. Having stored these results, we can then calculate the quantities of interest, the personal and partisan advantages, as before. As the table's final row shows, using this alternative technique does not produce results at variance with those from the other specifications and bandwidths.

Table A4. Robustness Of The Personal And Partisan Estimates To Alternative Specifications

			Vote share		Victory	
Polynomial	Both Sides	Bandwidth	Personal	Partisan	Personal	Partisan
1	No	.2	.083	−.006	.524	−.109
2	No	.2	.078	−.003	.530	−.114
3	No	.5	.092	−.023	.537	−.124
4	No	.5	.088	−.020	.533	−.121
1	Yes	.2	.080	−.004	.526	−.111
2	Yes	.2	.085	−.010	.753	−.300
3	Yes	.5	.041	.029	.679	−.242
4	Yes	.5	.087	−.017	.814	−.343
Kernel Weighted Local Polynomial			.078	−.008	.726	−.261

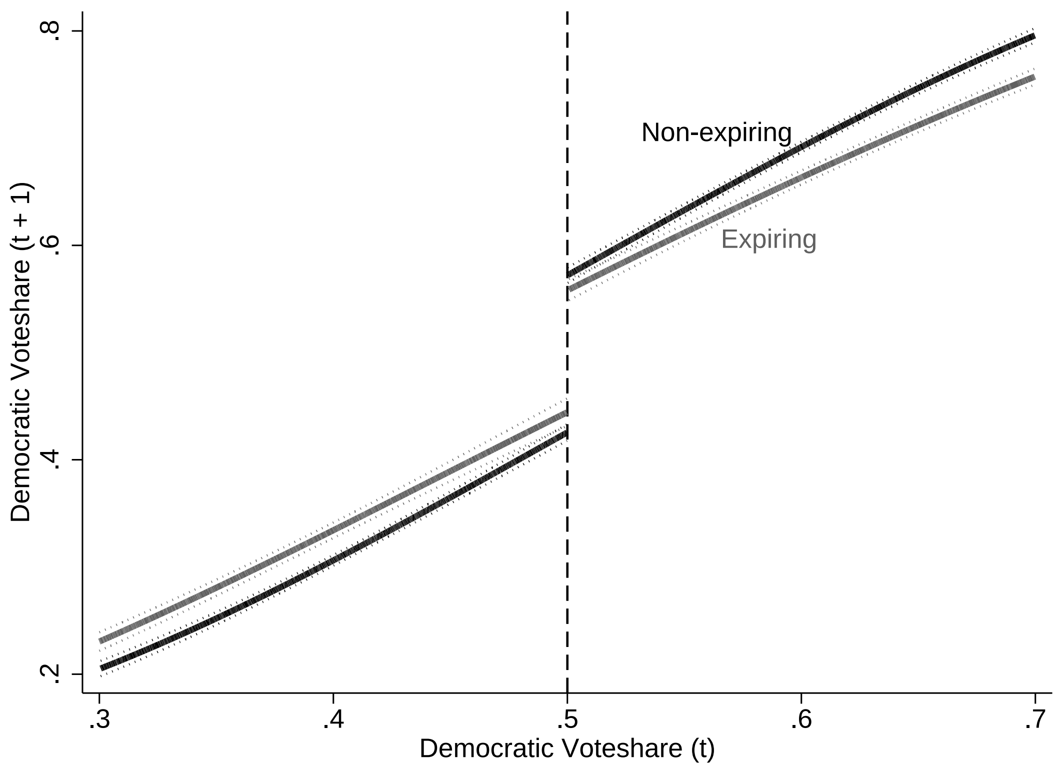
The table shows the robustness of our estimates to alternative specifications. We can vary the bandwidth, polynomial, and whether we separate models on each side of the threshold. In each case, the results are virtually unchanged.

3 Are Expiring Cases Meaningfully Different from Non-expiring Cases?

A critical assumption of our analysis is that the RD estimates would be the same in expiring and non-expiring cases if not for term limits forcing certain incumbents to retire. In order to explicitly test the validity of this assumption, we conduct a placebo where we focus on states without term limits and pretend that they adopted the term limit laws of California. We choose California because the state adopted term limits early, and their restrictions are similar to those in other state.

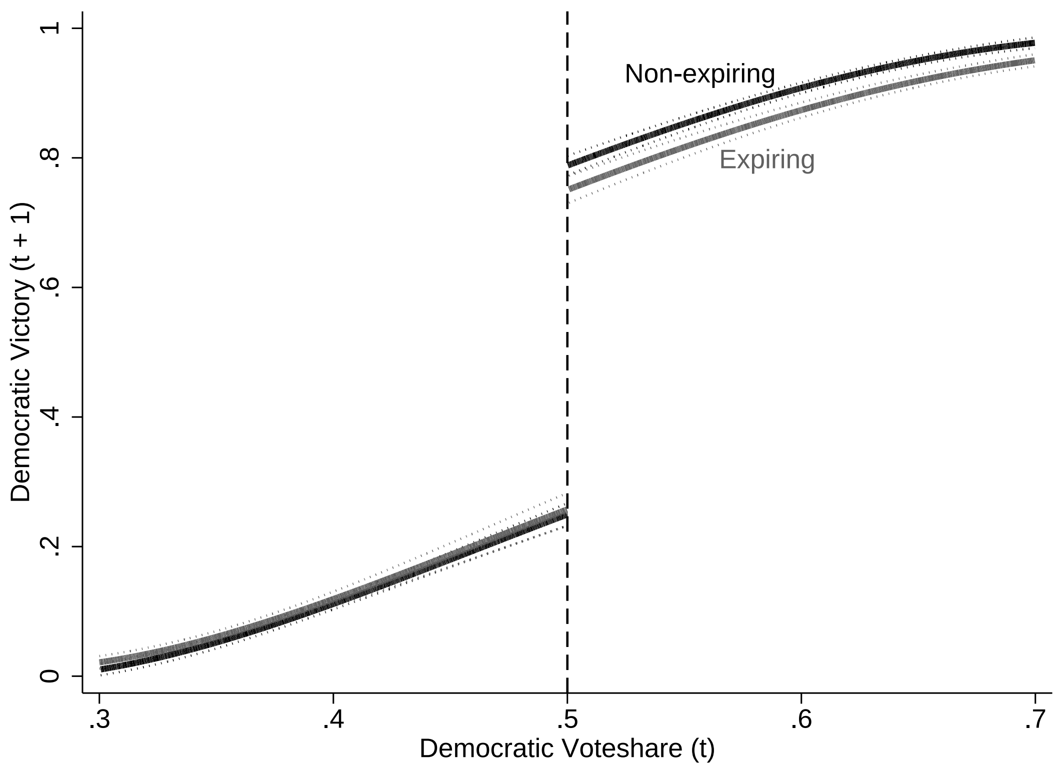
For each state without term limits, we can identify elections that would be classified as expiring or non-expiring if a California-style term limits scheme was present. We identify 2,233 elections where a candidate would have been expiring under CA's term limit laws and 3,630 elections that would have been non-expiring under CA's term limit laws. Then, we run our RD analysis to test whether we find a difference between these cases. If our assumptions hold, there will be no difference in the RD estimates for these elections where term limits were not in place. Indeed, this is what we find. Figures A1 and A2 replicate our analyses from Figures 2 and 3 for these cases without term limits. The RD estimates in the expiring and non-expiring cases are both large and statistically significant, as expected. However, they are substantively and statistically indistinguishable from one another. This suggests that expiring and non-expiring cases are otherwise comparable to one another and the only meaningful difference for our analysis is the fact that term limits force certain incumbents to retire in the expiring case.

Figure A1. Expiring and Non-expiring RD in States without Term Limits; Vote Share



The figure presents the RD estimates in placebo expiring and non-expiring cases, using states that do not have term limits laws. For each state, we pretend that it had adopted the same laws as California, and we find that the RD estimates are nearly identical, suggesting that expiring and non-expiring cases are comparable to one another.

Figure A2. Expiring and Non-expiring RD in States without Term Limits; Victory



The figure presents the RD estimates in placebo expiring and non-expiring cases, using states that do not have term limits laws. For each state, we pretend that it had adopted the same laws as California, and we find that the RD estimates are nearly identical, suggesting that expiring and non-expiring cases are comparable to one another.

4 Are Expiring and Non-expiring Elections Comparable?

One crucial assumption of our empirical strategy is that expiring and non-expiring cases are otherwise comparable if not for term limits sometimes forcing the winners of expiring elections out of office. In Table A5, we test whether expiring and non-expiring elections differ from one another in pre-treatment characteristics that may lead to biased estimates. We do so by regressing a pre-treatment, placebo outcome on an indicator variable for expiring cases along with state-chamber fixed effects. We find that expiring cases are no different than non-expiring cases in terms of prior partisanship or prior electoral competitiveness. Expiring cases are not more or less likely to have a Democratic vs. Republican incumbent, the previous election saw no more or less support for the Democratic Party, the Incumbent Party received no more or less support in the previous election, and the previous election was no more or less likely to have been uncontested. We also rerun these regressions including fourth-order polynomial controls for the Democratic vote share at time t. We include these controls, because our specific assumption is that expiring cases are comparable to non-expiring cases conditional on the present election's vote share. Again we find no meaningful differences between expiring and non-expiring cases. All point estimates are substantively negligible and statistically indistinguishable from zero.

We could also test whether the candidates running in expiring and non-expiring elections are comparable. Unfortunately, we are unable to collect detailed data about each state legislative candidate, but we do have information on one important variable–whether or not the winner of a close election seeks reelection. In Table 2, we find that the winner of a close election seeks reelection 68 percent of the time in non-expiring cases and 37 percent of the time in expiring cases. Half of the time, the term-limited candidate wins the close, expiring election. This means that candidates who are eligible to seek reelection do so about 7 out of 10 times regardless of whether they are running in expiring or non-expiring elections. The fact that winners seek reelection approximately half as often in expiring cases suggests that eligible winners seek reelection at the same rates in both expiring and nonexpiring cases, providing further evidence on the comparability of these cases.

5 Replicating Lee (2008) Estimates in Our Sample and Testing for Sorting

Here, we replicate the RD estimate of the incumbency advantage (Lee 2008) with our specific sample of state legislative elections and test for strategic sorting around the threshold. As discussed, RD estimation requires the researcher to make several decisions about the model specification. For each election in our sample, we identify the top performing Democrat and the top performing Republican. We exclude any race where a third party candidate

Table A5. Comparisons Between Expiring And Nonexpiring Elections Within Chambers

Placebo Outcome	DV Mean (Non-expiring)	Difference (Expiring − Non-expiring)	
		no controls	w/ controls
Dem. Vote Share (t-1)	.496	**.006**	**.000**
		(.009)	(.006)
Dem. Victory (t-1)	.454	**.007**	**−.005**
		(.018)	(.011)
Inc. Party Vote Share (t-1)	.678	**.008**	**.003**
		(.005)	(.005)
Uncontested Race (t-1)	.143	**.006**	**.002**
		(.012)	(.012)

Robust standard errors are in parentheses.
The table presents the results of 8 different regressions which compare expiring and non-expiring cases within each state-chamber with respect to 4 different pre-treatment variables. The estimates in the second column result from regressions of the placebo outcome on a binary indicator for expiring cases and state-chamber fixed effects. The results in the third column represent results from the same regression except that fourth-order polynomial controls are also included. In each case, the difference between expiring and non-expiring cases is substantively tiny and statistically indistinguishable from zero.

outperformed either of these candidates. Then we calculate the Democratic two-party vote share in the election at time t and in the election at time $t+1$ in that district.[1] These two variables represent the running variable and the dependent variable, respectively. We then specify a regression where the Democratic Party's subsequent vote share is a function of a fourth-order polynomial of the previous vote share and of a dummy treatment variable indicating whether the Democratic candidate won the previous election. We include all elections where the Democratic vote share at time t was between .4 and .6, but we do not assume that incumbency is randomly assigned for this entire range of data, a common misconception. Even though elections far from the threshold are included in the model for statistical efficiency, the model only estimates the discontinuity at the limit, so the assumption of "as if" random assignment is only necessary at the threshold. The coefficient on the treatment variable is the RD estimate, estimating the causal effect of the Democratic candidate barely winning the election compared to the candidate barely losing the election

[1] If a Democrat (Republican) runs unopposed at time $t+1$, we code her vote share as 1 (0), because she indeed garnered 100% (0%) of the two-party vote. Moreover, the ability to scare off challengers is one component of the incumbency advantage. However, our subsequent results are unchanged if we drop these observations. The coding of these cases is inconsequential because very few close elections are followed by uncontested races.

Table A6. RD Estimates In State Legislatures With Term Limits

	Vote share	Victory
Replication	.078	.372
	(.014)	(.048)
Placebo	.025	.028
	(.021)	(.052)
w/ Controls	.076	.371
	(.014)	(.047)
Donut RD	.069	.301
	(.018)	(.058)

Robust standard errors in parentheses.
The table replicates Lee's (2008) incumbency advantage estimate in state legislative elections with term limits. The sample includes all such elections where the Democratic candidate received between 40 and 60 percent of the vote. Results are comparable to those obtained by Lee in the U.S. House, suggesting the external validity of state legislatures. The replication shows that when the Democratic Party barely wins one election, they receive an extra 7.8 percentage points of the vote in the next election and are 37 percentage points more likely to win the next election compared to the scenario where they barely lose. The placebo tests show that prior election results do not predict the winner of close elections, relieving concerns about the validity of this method in this electoral setting. Rows 3 and 4 show that the RD estimate is unchanged when controlling for covariates or excluding extremely close elections where there exists a possibility of strategic sorting.

on her party's vote share or probability of victory in the following election. All of our results are robust to alternative specifications including different polynomial orders, different bandwidths, and local linear regression instead of polynomials.

Table A6 presents the results of this analysis. The RD estimate of the effect of Democratic incumbency on vote share is .078, indicating that a Democratic incumbent increases the party's vote share by 7.8 percentage points in the subsequent election. Similarly, Democratic incumbency increases the probability of a subsequent Democratic victory by 37 percentage points. These effects are substantively large, statistically significant ($p < .001$), and similar to Lee's estimates for the U.S. House.

Several recent studies express skepticism about the validity of RD in electoral settings (Caughey and Sekhon 2011; Grimmer et al. 2012). According to these studies, the winner of very close elections is not "as if" randomly assigned. Electoral fraud or resource imbalances could cause some candidates to selectively win more close elections than others. For example, incumbent candidates or candidates whose party aligns with the State Secretary appear to have an edge in very close Congressional elections. We are sensitive to these concerns, and we present several tests to address these possibilities. Each test indicates that any non-random sorting around the threshold, if it exists, does not pose a problem for our subsequent results.

The second row of Table A6 conduct placebo tests of the RD estimate where the dependent variable is replaced with the Democratic Party's vote share or victory in the previous election. If incumbents are better at winning very close elections, then we would see large, positive estimates. However, our results suggest that the incumbent party is not better at winning close elections than the non-incumbent party. Also in Table A6, we present two robustness tests which further address these challenges. In Row 3, we include two additional controls, lag vote share and lag incumbency, to ensure that slight imbalances in these variables do not bias our result. In Row 4, we conduct "donut" RD analyses (Almond and Doyle 2011; Barreca et al. 2011; Barreca, Lindo, and Waddell 2011) where we exclude all elections where one candidate won by less than half a percentage point. If any non-random sorting does occur in these elections as purported by Caughey and Sekhon (2011), it will occur for extremely close elections where fraud, recounts, court cases, or other non-random factors could push one candidate over the top. By excluding these cases, the identifying power of RD still holds but concerns about strategic sorting have been mitigated. The RD estimates are virtually unchanged with both of these robustness checks.

The RD estimate of the incumbency advantage is highly robust and strikingly similar to that in the U.S. House, suggesting that state legislative elections with term limits may provide a useful laboratory for learning about the incumbency advantage in general.

We also test for sorting using the technique proposed in McCrary (2008), as implemented in the Stata .ado file DCDensity.ado (available at http://eml.berkeley.edu/~jmccrary/DCdensity/). This technique tests for differences in the density of a variable on each side of the discontinuity. We apply it to two variables: the main running variable, which is the Democratic vote share winning margin, and the incumbent party vote share winning margin. We also test for these discontinuities in three settings: in all elections, in only expiring cases, and in only non-expiring cases. In all cases we find no evidence for discontinuities.

Figure A3 presents these tests graphically, and Table A7 presents the results of formal statistical tests. Each figure plots the density of the variable of interest separately on each side of the discontinuity. We see no visual evidence of sorting in any case, and we fail to reject the null hypothesis of no sorting in all cases.

Figure A3. McCrary Density Tests for Democratic and Incumbent Win Margin

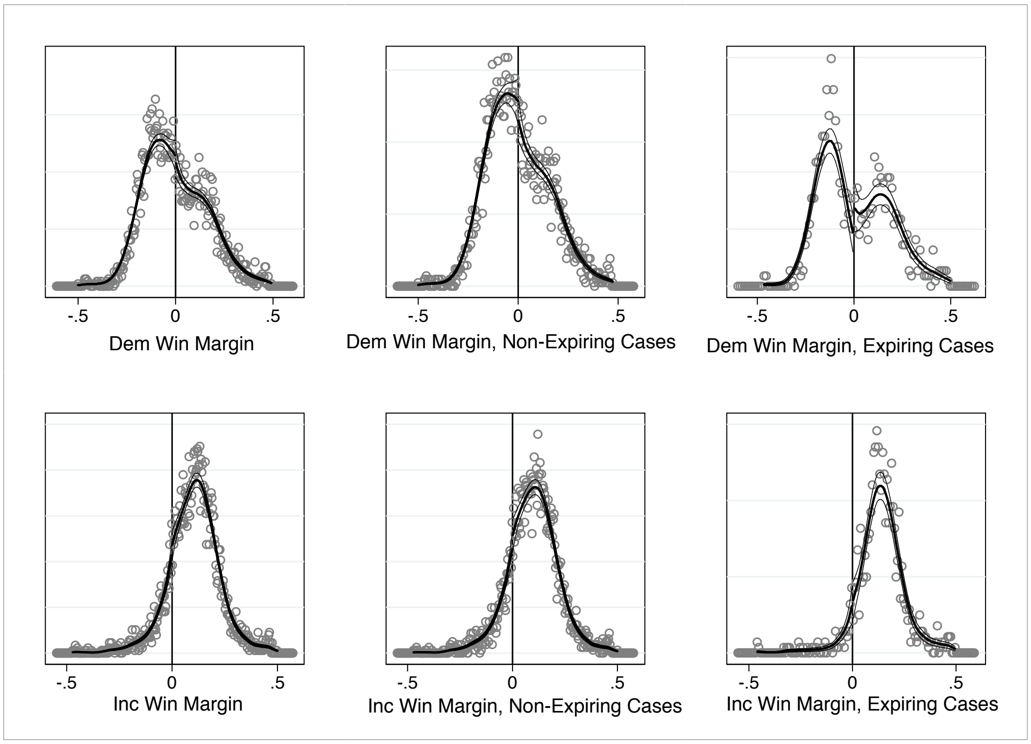


Table A7. McCrary Tests for Sorting

	Full Sample	Non-Expiring	Expiring
Dem Win Margin	-.046 (.085)	-.097 (.092)	.469 (.270)
Inc Win Margin	.064 (.102)	.090 (.107)	.001 (.287)

Standard errors in parentheses.
The McCrary test statistic is the discontinuity in log density around the electoral threshold. We find no evidence of sorting or clumping in any case.

6 Comparison to Other Estimates of the Incumbency Advantage

Table A8. Comparison with Previous Estimators using the Same Data

Estimator	Similar to . . .	St. Leg.	U.S. House
Sophomore Surge	Erikson 1971	.067	.041
Retirement Slump	Alford and Brady 1989	.032	.105
Panel Regression	Ansolabehere and Snyder 2002	.059	.085
Regression Discontinuity	Lee 2008	.078	.058
RD + Term Limits; Personal	This Study	**.088**	NA
RD + Term Limits; Partisan		**−.020**	

The table presents five previous estimates of the incumbency advantage for state legislative elections with term limits and U.S. House elections from 1994 to 2008, respectively. The last 2 rows present the personal and partisan incumbency advantages as estimated in this study.

We benchmark our estimates against previous estimators for the same set of elections. In this section, we focus exclusively on the vote share estimates which have been the predominant focus of previous literature. Table A8 presents five previous estimates of the incumbency advantage using the same data set employed throughout this paper - state legislative elections with term limits - along with U.S. House elections for a similar period, 1994 to 2008.

The sophomore surge (Erikson 1971) is .067 and .041 for state legislative elections and House elections, respectively. Newly-elected legislators improve by 4 to 7 percentage points in their next election (if they seek reelection). The retirement slump (Alford and Brady 1989) is .032 and .105 for these two sets of elections, indicating that a party loses 3 to 11 percentage points when its incumbent retires. As expected, the retirement slump is significantly smaller in state legislative elections with term limits than in the U.S. House, because many retirement decisions are forced by term limits in the former case, while retirements in the House may be more strategic. Row 3 of Table A8 presents a panel regression estimate of the incumbency advantage in the spirit of Gelman and King (1990) or Ansolabehere and Snyder (2002). We regress the Democratic Party's vote share in a given election on an incumbency variable (1 = Democratic incumbent, 0 = open seat, −1 = Republican incumbent), district-decade fixed effects, and chamber-year fixed effects. The resulting estimates of .059 and .085 indicate that a party tends to perform 6 to 9 percentage points better with an incumbent running than in an open seat. Again, the estimate is significantly smaller in state legislatures with term limits than in the House, and again this could be plausibly attributed to the presence of more strategic retirements in the House.

Next, we present the traditional RD estimate (Lee 2008) in both settings. For U.S. House elections from 1994 to 2008, the RD estimate is .058, similar to the .078 that we already estimated for state legislative elections.

Last, we present our own estimates of the personal and partisan incumbency advantages at the bottom of Table A8. No other approach makes a distinction between the personal and partisan incumbency advantages. Our results suggest that, to the extent that previous studies detect an incumbency advantage, this advantage is largely personal.

7 Discussion of Competitive Elections and our Local Average Estimates

Because our empirical design draws inferences from close elections, our estimates are local to those electoral settings. As we describe in the main text, we estimate the personal and partisan incumbency advantages for the hypothetical set of settings where the last election was a 50-50 tie between a Democrat and Republican. On one hand, our local estimates are a limitation of our design and a necessary cost that we must pay in exchange for weaker identifying assumptions. On the other hand, however, competitive electoral settings are of great substantive and political interest, and they are precisely those for which we would most like to understand the effects of incumbency.

Despite the importance of competitive electoral settings, we might want to say something about the effects of incumbency in other settings as well. As discussed above, if the effects of personal or partisan incumbency are notably different in uncompetitive settings or highly partisan electorates, then our local estimates could diverge from the average effects. However, there is no obvious theoretical reason to expect that the effects of personal or partisan incumbency will be greater or smaller in uncompetitive settings. Explanations in either direction are easy to construct. Candidates, voters, and parties will behave differently in competitive settings which could produce greater advantages if, for example, candidates and parties better exploit the benefits of incumbency in close races, or smaller advantages if, for example, voters select more on candidate quality and other factors unrelated to incumbency in close races. Hainmueller, Hall, and Snyder (2014) surveyed 165 political scientists, asking them to guess whether incumbency effects were greater or smaller in less competitive electoral settings. Interestingly, scholars were divided on the question with about one-third expecting them to be larger, one-third expecting them to be smaller, and one-third expecting them to be about the same. Hainmueller, Hall, and Snyder go on to estimate incumbency effects in less competitive settings and they find little evidence of heterogeneity. Therefore, there is

Table A9. Personal And Partisan Advantages In Relatively Moderate Or Partisan Districts

	Personal	Partisan
Moderate Districts	.097	−.025
	[−.045, .239]	[−.113, .066]
Partisan Districts	.100	−.013
	[−.041, .319]	[−.151, .073]

Bootstrapped 95% Confidence Intervals in Brackets.
The personal and partisan incumbency advantages do not appear to differ across moderate and partisan districts.

little scholarly consensus or empirical evidence on the heterogeneity of incumbency effects in less competitive settings. Of course, when we examine the probability of victory as the dependent variable, we do expect that the local average effects of incumbency will be greater than the average effects across all cases, because the same effect in terms of vote share is more likely to influence the election result in close settings. However, when examining vote share as the outcome, there is little reason to expect significant heterogeneity.

Although we draw inferences from 50-50 elections, the sample of districts that sometimes have close elections is more representative than one might expect. In our sample, 37.3 percent of all state legislative districts elected both a Democrat and a Republican at some point in our period of study, so a significant portion of districts could potentially have a close race. Moreover, when we look at the sample of districts in our sample that did have a close race (within 2 percent) at some point, the average two-party Democratic vote share ranges from 35.2 to 68.4 percent. In other words, highly partisan districts do have close elections from time to time, and these cases contribute to our estimates. Therefore, our local average estimates may not be as local as one might have expected. A significant proportion of electoral settings, some of which are highly partisan, have close elections.

Finally, we explicitly test for heterogeneity by separately estimating the personal and partisan incumbency advantages in relatively moderate districts where the average two-party vote share falls between 45 and 55 percent and in relatively partisan districts where the average two-party vote share is less than 45 or greater than 55 percent. In splitting the data, we lose statistical power, and the subsequent results are imprecise. However, because partisan districts do have close races from time to time, we can estimate the personal and partisan incumbency advantages for these cases. Table A9 presents the results. The estimates, while imprecise, are nearly identical for moderate and partisan districts, suggesting that our local estimates may not be significantly unrepresentative of the average effects in a broader set of districts.

8 Considering Bias from Replacement Candidate Quality

Assumption 3 states that when an incumbent retires after narrowly winning a close election, the new candidate running from the same party is, on average, of comparable quality. Let us define QD as the average difference in the quality of a retiring incumbent who just won her last election and the quality of the new candidate from the same party who runs in the next open-seat election.

Define QD such that positive values indicate that the new candidate is higher quality than the retiring incumbent and negative values indicate that the new candidate is lower quality. Imagine that we could calculate the average quality differential for Democrats (QD_D) and the average quality differential for Republicans (QD_R). Our assumption requires that the average (or sum) of these two quantities is zero. In a Democratic-leaning district, we would expect close elections to result from low-quality Democrats, high-quality Republicans, or some combination of the two, so we would expect that $QD_D > 0$ and $QD_R < 0$ in such a case. However, this is not a violation of our assumptions so long as $QD_D + QD_R = 0$. Later in this section, we show analytically that such a condition will hold even in partisan districts and even when random errors influence the election result. Since Democrats and Republicans win extremely close elections at equal rates, QD (the average quality difference across all cases of an incumbent retirement after a close election) simply equals the average of QD_D and QD_R.

If we allow for non-zero values of QD, then Lee's RD design estimates the function

$$RD = 2 \cdot Personal \cdot P + 2 \cdot Partisan + 2 \cdot QD \cdot (1 - P). \tag{1}$$

Therefore, when we implement the RD design in expiring and non-expiring cases, we estimate the two functions:

$$RD_1 = 2 \cdot Personal \cdot P_1 + 2 \cdot Partisan + 2 \cdot QD \cdot (1 - P_1), \tag{2}$$

and

$$RD_0 = 2 \cdot Personal \cdot P_0 + 2 \cdot Partisan + 2 \cdot QD \cdot (1 - P_0). \tag{3}$$

We estimate the personal incumbency advantage through the equation

$$\hat{Personal} = \frac{RD_0 - RD_1}{2(P_0 - P_1)}. \tag{4}$$

Substituting from above, we get

$$\begin{aligned} \hat{Personal} &= \frac{2PersonalP_0 + 2Partisan + 2QD(1-P_1) - 2PersonalP_1 - 2Partisan - QD(1-P_0)}{2(P_0 - P_1)} \\ &= Personal - QD. \end{aligned}$$

We estimate the partisan incumbency advantage through the equation

$$\hat{Partisan} = \frac{RD_1 P_0 - RD_0 P_1}{2(P_0 - P_1)}. \tag{5}$$

Substituting from above, we get $\hat{Partisan} =$

$$\frac{([2PersonalP_1 + 2Partisan + 2QD(1-P_1)]P_0 - [2PersonalP_0 + 2Partisan + 2QD(1-P_0)]P_1)}{2(P_0 - P_1)}$$
$$= Partisan + QD.$$

Therefore, if $QD \neq 0$, our estimates of the personal and partisan incumbency advantages will be biased. In particular, we will underestimate the personal advantage by QD and we will overestimate the partisan advantage by QD. If new candidates are systematically worse than retiring incumbents who barely won their last election ($QD < 0$), then we will overestimate the personal advantage and underestimate the partisan advantage. With these results, we can assess the sensitivity of our empirical results to different values of QD. In Table 2, we estimate the personal and partisan advantages as .088 and -.020, respectively. If there were no personal incumbency advantage, replacement candidates would have to be 8.8 percentage points worse, on average, than retiring incumbents who barely won their last election in order for us to obtain the estimates that we do. Similarly, suppose that the personal and partisan advantages were equal in their magnitudes. Replacement candidates would have to be 5.4 percentage points worse, on average, for us to obtain the estimates that we do. Both of these hypothetical values of QD seem implausible. The average difference in the quality of replacement candidates and retiring incumbents who barely won their last election seems likely small to negligible, and therefore the resulting bias of our estimates is also likely small to negligible. Below, we provide a simple model which explains why we might expect QD to be zero or close to zero.

8.1 Model of Candidate Quality and Electoral Selection

We now turn to a simple model of candidate quality and electoral outcomes to investigate the expected average quality of candidates at the RD threshold. After showing that, under the simple model we present, average candidate quality at the threshold is equal to zero, we also consider perturbations that would lead to violations of this assumption.

Democratic and Republican candidates have "quality," thought of in terms of bonus vote share, equal to Q_D and Q_R, respectively. These qualities are distributed as $Q_D \sim N(0, \sigma_p^2)$ and $Q_R \sim N(0, \sigma_p^2)$. The district has a fixed Democratic normal vote, N, benchmarked to the Democratic vote-share winning margin, V_D. The Democratic candidate wins the election with $V_D > 0$. Finally, the district is subject to a shock, $\epsilon \sim N(0, \sigma_\epsilon^2)$.

Therefore the Democratic vote share in the district is

$$V_D = N + Q_D - Q_R + \epsilon, \tag{6}$$

or, rearranging and defining $\nu = -Q_R + \epsilon$,

$$V_D - N = Q_D + \nu. \tag{7}$$

Interest is in calculating the expected quality of the Democratic and Republican candidates, conditional on the observed vote share winning margin (and especially conditional on this vote share margin being 0, as in the RD).

Consider the expected Democratic quality after observing the election outcome, $V_D - N$. The "prior" mean for Q_D is zero, given by assumption above. Our maximum likelihood estimate for Q_D after observing $V_D - N$ is the value that maximizes $P(V_D - N | Q_D)$, which is clearly $V_D - N$ since $E[\nu] = 0$. Finally, note that $Var(V_D - N | Q_D) = Var(Q_D + \nu | Q_D) = Var(\nu) = \sigma_p^2 + \sigma_\epsilon^2$.

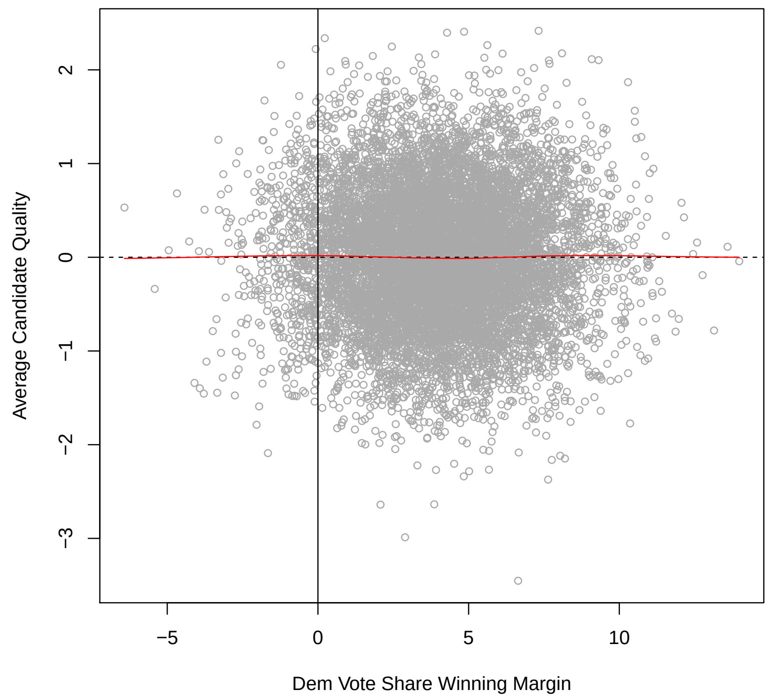
Because of the normality of the variables, we can use the well-known result that the posterior mean is the precision-weighted mean of these two quantities (see for example De Groot 2004: 167). Therefore we have

$$E[Q_D | V_D - N] = \frac{0 \cdot \frac{1}{\sigma_p^2} + (V_D - N) \cdot \frac{1}{\sigma_p^2 + \sigma_\epsilon^2}}{\frac{1}{\sigma_p^2} + \frac{1}{\sigma_p^2 + \sigma_\epsilon^2}},$$

Or, simplifying and conditioning on the RD cutoff of $V_D = 0$,

$$E[Q_D | V_D - N] = \frac{-N}{2 + \frac{\sigma_\epsilon^2}{\sigma_p^2}}. \tag{8}$$

Figure A4. Simulation with Equal Variances



We can apply the exact same logic for Republicans. By symmetry, it is easy to show that

$$E[Q_R|V_D - N] = \frac{N}{2 + \frac{\sigma_\epsilon^2}{\sigma_p^2}}. \tag{9}$$

Therefore, $E[Q_R + Q_D|V_D - N] = 0$, and thus the average candidate quality is zero.

A simple simulation illustrates this in practice. We set $\sigma_D^2 = \sigma_R^2 = 1$ and $\sigma_\epsilon^2 = 2$, with $N = 4$. Over 10,000 iterations we draw Q_D, Q_R, and ϵ from these distributions and calculate the resulting V_D. In Figure A4, we plot the resulting V_D on the horizontal axis against the average candidate quality—that is, the Democratic and Republican qualities in the election summed and divided by 2—in the election, on the vertical axis. The vertical line indicates the RD threshold, when the election is exactly tied. The red line is a loess smoother measuring the conditional mean of average candidate quality for levels of the Dem vote share winning margin. As we see, this conditional mean is 0 at the RD threshold.

Thus we see that, under this simple model of candidate quality and electoral outcomes, average candidate quality at the RD threshold will be zero. Though in a Democratic-leaning district the Democrat might be of lower than average quality and/or the Republican might be of higher average quality, these cancel out such that, on average, the average of the two is zero. When the winners of these races are then forced to retire in the next election cycle, the replacement candidate, drawn from these same distributions in the model, will be no higher or lower in quality, on average.

8.2 Perturbing the Model to Obtain Non-Zero Average Quality at the RD Threshold

It is important to consider how robust this simple model's conclusions are. In particular, we show here that candidate quality will no longer be zero at the RD threshold if the variances in the parties' candidate quality distribution are different. While this is important to consider, we struggle to identify a theoretical justification for why these variances would differ.

We start from the model as before, but with $Q_D \sim N(0, \sigma_D^2)$ and $Q_R \sim N(0, \sigma_R^2)$.

A bit of algebra establishes that

$$E[Q_D|V_D = 0] = \frac{-N}{\left(\frac{1}{\sigma_D^2} + \frac{1}{\sigma_R^2+\sigma_\epsilon^2}\right)(\sigma_R^2 + \sigma_\epsilon^2)} \tag{10}$$

and

$$E[Q_R|V_D = 0] = \frac{N}{\left(\frac{1}{\sigma_R^2} + \frac{1}{\sigma_D^2+\sigma_\epsilon^2}\right)(\sigma_D^2 + \sigma_\epsilon^2)}. \tag{11}$$

For average quality to be zero we must have $E[Q_D|V_D = 0] + E[Q_R|V_D = 0] = 0$ and therefore $E[Q_D|V_D = 0] = -E[Q_R|V_D = 0]$. Thus we must have

$$\frac{N}{\left(\frac{1}{\sigma_D^2} + \frac{1}{\sigma_R^2+\sigma_\epsilon^2}\right)(\sigma_R^2 + \sigma_\epsilon^2)} = \frac{N}{\left(\frac{1}{\sigma_R^2} + \frac{1}{\sigma_D^2+\sigma_\epsilon^2}\right)(\sigma_D^2 + \sigma_\epsilon^2)}.$$

Simplifying, this condition becomes

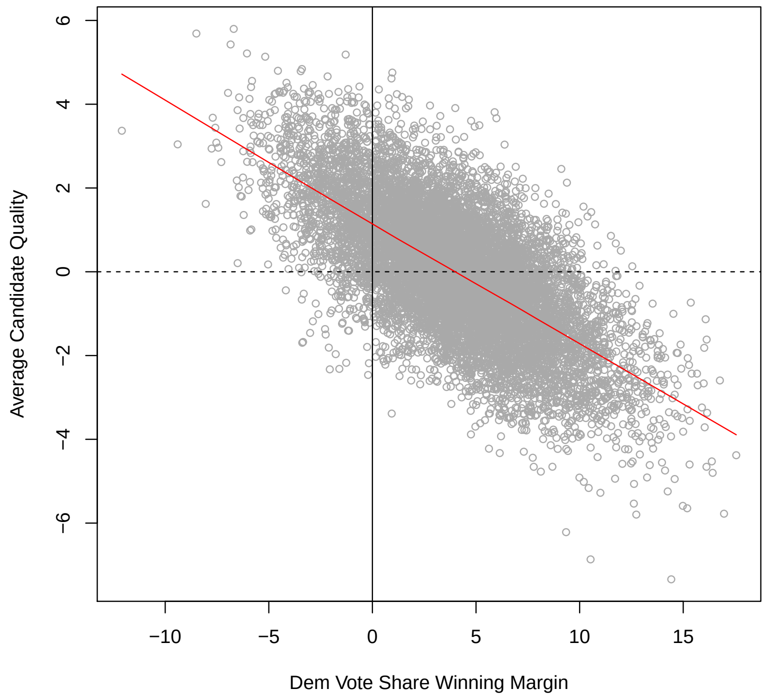
$$\frac{\sigma_D^2 N}{\sigma_R^2 + \sigma_\epsilon^2 + \sigma_D^2} = \frac{\sigma_R^2 N}{\sigma_D^2 + \sigma_\epsilon^2 + \sigma_R^2}.$$

Thus we must have $\sigma_D^2 N = \sigma_R^2 N$, and therefore it must be the case that $\sigma_D^2 = \sigma_R^2$.

To illustrate this phenomenon by simulation, we repeat the simulation from above except we set $\sigma_R^2 = 3$ so that $\sigma_R^2 \neq \sigma_D^2$. As the Figure A5 shows, we no longer have the conditional mean of average candidate quality equal to 0 at the discontinuity (where the red loess line intersects the vertical line).

It is difficult to calibrate any such model to actual data since we cannot observe candidate quality. Nevertheless, we think it is unlikely that the variances in the party's quality distributions could be large enough—and different enough—to account for the estimates we have

Figure A5. Simulation with Unequal Variances and Systematic Advantage to One Party



observed. Under the reasonable assumption that candidate quality is distributed normally and identically across parties, the quality differential QD we discussed above will be zero and our estimates will include no bias from replacement candidate quality differences. In addition, as we show in the next section, under a reasonable assumption about how the normal vote changes across districts, we can recover average candidate quality at the discontinuity even with unequal variances in candidate quality across the two parties.

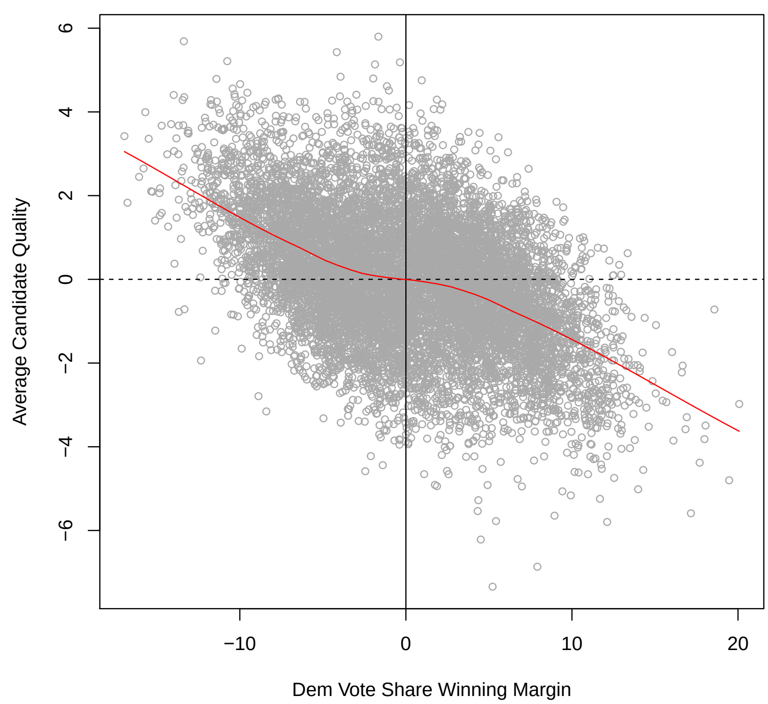
8.3 Equal Candidate Quality with Unequal Variances

A final tweak to the model allows for equal candidate quality even when the party variances are unequal. Instead of treating N as a fixed quantity in a single district (e.g., always advantaging Democrats), we now imagine looking across many districts with different normal votes. Intuitively, we should expect this change to recover equal candidate quality at the discontinuity if the distribution of N is symmetric around zero. Although candidate quality will be unequal in the Democratic-advantaged districts and in the Republican-advantaged districts at the RD threshold, as in the previous section, the two will cancel each other out under this condition.

To illustrate this, we run a final simulation in which we draw N from a mixture of normal distributions. First, we draw $B \sim Bernoulli(0.5)$. Next, we generate 10,000 draws from $N_D \sim N(4, 1^2)$ and 10,000 draws from $N_R \sim N(-4, 1^2)$. Finally, we construct $N = B \cdot N_D + (1 - B) \cdot N_R$, and we calculate the resulting vote shares and average candidate qualities as before. As Figure A5 shows, we now have average candidate quality equal to zero conditional on the winning margin equaling zero.

This tweak closely fits our observed data. Some state legislative districts lean Democratic, some lean Republican, but the distribution of normal two-party vote shares is fairly symmetrical around the 50-50 threshold. Among state legislative districts that at some point had a close election (within 2 percent) in our sample, the average two-party Democratic vote share is 49.9 percent. In other words, because neither party has a systematic advantage across state legislative districts, we expect the average quality of the winners of close elections to be average, regardless of whether the variances in quality differ between the parties.

Figure A5. Simulation with Unequal Variances and No Systematic Advantage to One Party



8.4 Incumbent Quality in Close Elections

A final concern warrants further discussion. Our use of term limits means that the close races in expiring cases feature at least one incumbent in a close election at time t. The fact that an incumbent is drawn into a close election suggests that she may be of lower than average quality, which in turn might mean that the replacement candidate is of higher quality than the termed-out incumbent, despite the modeling exercises presented above. On the other hand, acting in the opposite direction is the fact that she has previously won election to office when she was not the incumbent, making her more likely to be of higher quality. Depending on the level of noise in the election results, these effects could cancel out or either one might outweigh the other.

Thus in determining whether a candidate is likely to be of lower or higher than average quality, conditional on being in a close election at time t as an incumbent, we must consider whether the prior incumbent victories are more or less informative about her quality than the current close election. It is difficult to theorize precisely over how this dynamic will play out, making it possible that replacement candidates are either below or above the average quality of their retiring incumbents, on average. However, several possibilities would render this question irrelevant. First, if the distribution of candidate quality is relatively tight, then the incumbents will not differ much from their replacements in either case. Second, if the return to quality in terms of vote share is low, then again any remaining differences will be unimportant.

Although we cannot test for this threat directly, since candidate quality is unobservable, we can carry out an indirect test. If the results are driven by the violation of the replacement candidate quality assumption, then the estimate should differ across cases in which the incumbents are of higher or lower quality, on average. Incumbents who have served for longer when they are forced out by term limits are likely to be of higher average quality than incumbents who are forced out after serving fewer terms. Therefore, we investigate whether the estimated effects differ when we exclude the first election after term limits are put in place. Dropping these observations means excluding the longest-serving incumbents, the first ones forced out by term limits, and thus should lower the average quality of the retiring incumbents left in the sample. As Table A10 shows, however, we continue to find extremely similar estimates for the personal and partisan incumbency advantages when we exclude these cases. As a result, any differences in the average quality of term-limited incumbents and their replacement candidates is unlikely to drive our estimates.

Table A10. Estimating the Personal and Partisan Incumbency Advantages after Excluding the First Eligible Election for each State-Chamber

	Non-expiring	Expiring	Difference
RD Estimate (Vote Share)	.080	.010	.070*
	[.057,.103]	[−.049,.069]	[.005,.130]
RD Estimate (Victory)	.469	.181	.289
	[.416,.521]	[.004,.353]	[.104,.467]
Pr(Winner Runs Again)	.660	.297	.363
	[.606,.714]	[.154,.458]	[.194,.524]
Observations	3969	698	
Personal Advantage (Vote Share)		**.096**	
		[.007,.201]	
Partisan Advantage (Vote Share)		**−.024**	
		[−.091,.032]	
Personal Advantage (Victory)		**.397**	
		[.173,.763]	
Partisan Advantage (Victory)		**−.027**	
		[−.248,.121]	

Bootstrapped 95% Confidence Intervals in Brackets.
The table replicates Table 2, but excludes the first eligible year for each state-chamber. The first year where an election could be classified as expiring represents where we might be particularly worried that replacement candidates will be lower quality than the term limited incumbents. The fact that these results line up closely with the results in Table 2 suggests that differential candidate quality does not bias our estimates.

Appendix References

Calonica, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. Robust Data-Driven Inference in the Regression-Discontinuity Design. *The Stata Journal*, forthcoming.

DeGroot, Morris H. 2004. Optimal Statistical Decisions. Wiley and Sons, Inc.

Hainmueller, Jens, Andrew B. Hall, and James M. Snyder, Jr. 2014. Assessing the External Validity of Election RD Estimates: An Investigation of the Incumbency Advantage. Working Paper.

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