MIT
POLITICAL SCIENCE

Massachusetts Institute of Technology
Political Science Department

Working Paper No. 2013-26

On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from Over 40,000 Close Races

Andrew Eggers, London School of Economics
Olle Folke, Columbia University
Anthony Fowler, University of Chicago
Jens Hainmueller, MIT
Andrew B. Hall, Harvard University
James M. Snyder Jr., Harvard University

On The Validity Of The Regression Discontinuity Design For Estimating Electoral Effects: New Evidence From Over 40,000 Close Races

Andrew C. Eggers – London School of Economics
Olle Folke – Columbia University
Anthony Fowler – University of Chicago
Jens Hainmueller – Massachusetts Institute of Technology
Andrew B. Hall – Harvard University
James M. Snyder, Jr. – Harvard University and NBER

First Version: May 2013
This version: June 2013

Abstract

Many papers use regression discontinuity (RD) designs that exploit "close" election outcomes in order to identify the effects of election results on various political and economic outcomes of interest. Several recent papers critique the use of RD designs based on close elections because of the potential for imbalance near the threshold that distinguishes winners from losers. In particular, for U.S. House elections during the post-war period, lagged variables such as incumbency status and previous vote share are significantly correlated with victory even in very close elections. This type of sorting naturally raises doubts about the key RD assumption that the assignment of treatment around the threshold is quasi-random. In this paper, we examine whether similar sorting occurs in other electoral settings, including the U.S. House in other time periods, statewide, state legislative, and mayoral races in the U.S., and national and/or local elections in a variety of other countries, including the U.K., Canada, Germany, France, Australia, India, and Brazil. No other case exhibits sorting. Evidently, the U.S. House during the post-war period is an anomaly.

For generously providing data, the authors thank Melissa Dell, Fernando Ferreira, Alexander Fouirnaies, Ronny Freier, Danny Hidalgo, Yusaku Horiuchi, and Carl Klarner. For helpful comments, we thank Devin Caughey, Justin Grimmer, Gary King, and Jas Sekhon.

I. Introduction

In recent years the regression discontinuity (RD) design has become a workhorse model for causal inference in electoral research with numerous studies exploiting "close" elections to estimate the effects of election results on various political and economic outcomes of interest.[1] The RD has quickly become a popular empirical strategy because it offers an unusual opportunity to disentangle the effects of elections from factors that influence elections, at least for the subset of elections that are decided very narrowly. Lee (2008) formalizes the logic of the RD design based on close elections, and gives precise conditions under which the outcome of close elections can be used as a quasi-random treatment variable.

Three important recent papers criticize election-based RD studies because the electoral outcomes sometimes exhibit substantial imbalance near the threshold that distinguishes winners from losers. That is, observable attributes of one of the candidates - such as incumbency status, or whether the candidate has the same party affiliation as the officials who are presumed to control key features of the electoral process - appear to be significantly correlated with victory even in very close elections. Jason Snyder (2005) shows that in U.S. House elections between 1926 and 1992, incumbents win noticeably more than 50% of the very close races. Caughey and Sekhon (2011) investigate this further, and show among other things that winners in close U.S. House races raise and spend more campaign money. Grimmer et al. (2012) show that U.S. House candidates from the party in control of state offices, such as the governorship, secretary of state, or a majority in the state house or state senate, hold a systematic advantage in close elections.[2] These papers argue that the observed imbalances are evidence of "strategic sorting" around the election threshold. The fact that certain types of candidates appear to

[1]Examples include 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 Jr. (2012), and Gagliarducci and Paserman (2012).

[2]We are also aware of one other working paper identifying a potential concern with the RD design in close elections. Vogl (2012) finds that black candidates are better at winning close races than their white opponents in mayoral races in the U.S. South (but not elsewhere). However, from a statistical standpoint, the evidence is rather weak as there have been very few close mayoral races in the South between a white and black candidate. In Vogl's sample, there are only 38 such cases (from 18 unique cities) where the margin of victory was less than 20 points.

have a systematic advantage in tight races suggests a normatively substantial problem with the administration of elections. Somewhat more prosaically, this type of sorting naturally raises doubts about the prospect of using the RD design to learn about politics, given that it violates the key RD assumption that the assignment of treatment for an arbitrarily small window around the threshold is quasi-random. Under quasi-random assignment, for any confounder X_i, near the threshold we expect $Pr(Win_i \mid X_i) = 0.5$.

In this paper, we discuss the existing evidence and theory for sorting in recent U.S. House races and examine whether similar sorting occurs in other electoral settings. We study the U.S. House in other time periods, as well as statewide, state legislative, and mayoral races in the U.S. We also study national and/or local elections in a variety of other countries, including the U.K., Canada, Germany, France, Australia, India, and Brazil. We do not find a single other case that exhibits sorting. We conclude that the post-war U.S. House is an anomalous case that does not pose a general threat to the validity of RD designs in electoral settings.

We conclude the paper by providing recommendations to future researchers estimating electoral effects using RD designs. Consistent with Caughey and Sekhon (2011), we argue that the burden is on empirical researchers to justify their assumptions with theory and data. Moreover, statistical imbalance at an electoral threshold is expected to arise by chance from time to time and does not automatically invalidate the underlying assumption of an RD design. Nonetheless, applied researchers should be careful and thorough in showing that their results are not plagued by imbalance or misspecification. In short, the RD design is a powerful tool for estimating electoral effects and the discovery of imbalance at approximately the rate expected by chance should not discourage its use or acceptance among researchers.

II. Problematic close elections

The appeal of the RD design in the analysis of elections derives from the idea that the winner of a very close election is determined as if by a coin flip. In a race in which the winner and loser are separated by just a few votes, it seems plausible that the outcome could easily have been reversed if, for example, the weather on election day had been different. The quasi-random determination of who wins means that, in a large sample of close elections, the winners and losers should be similar on average and so should the contexts in which one type of candidate (e.g. the Democrat) wins and loses. This fundamental comparability across settings with different election outcomes offers an unusual opportunity to study the effects of these election outcomes on policy, representation, and other phenomena of interest.

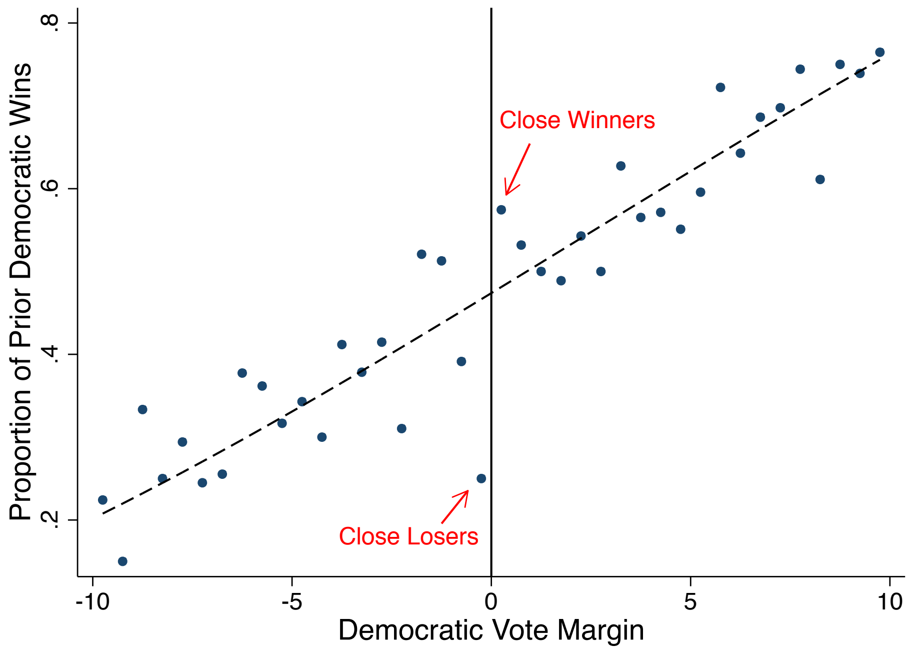
As noted above, however, evidence has accumulated to suggest that winners and losers are not in fact comparable in close elections to the U.S. House of Representatives; winners of close elections appear to be disproportionately incumbents (Snyder 2005) who are aligned with the locally-dominant party (Grimmer et al. 2012) and, among other differences, have more experience and money (Caughey and Sekhon 2011). These imbalances appear to pose a threat to the validity of the RD design: close elections are in principle coin flips, but the coin appears to be weighted.[3]

Figure 1 offers one view of the problem in the U.S. House of Representatives for the period 1946-2010. For each 0.5 point bin of Democratic vote margin (for example, all elections where the Democrat won by between .5% and 1% of overall vote share), we plot the proportion of cases in which a Democrat won the district in the previous election. Not surprisingly, the proportion of cases in which a Democrat was the incumbent is positively correlated with the Democratic vote margin and is at about .5 in close races where the margin is near 0. What stands out in the figure is the gap between the value for the bins immediately on either side of 0: in cases in which the Democrat *won* by less than .5% (i.e. the first bin to the right of the threshold), a Democrat previously won the seat almost 60% of the time; in cases in which the Democrat *lost* by less than .5% (i.e. the first bin to the left of the threshold), a Democrat previously won the seat only about 25% of the time. Given that both sets of elections are extremely close, we would expect the incumbent party to win the seat with probability approximately 1/2, but in fact the incumbent party seems to do much better than that in these very close elections. This highlights the anomaly first identified by Snyder (2005) and pursued further by Grimmer et al. (2012) and Caughey and Sekhon (2011).

What accounts for the disproportionate success of the incumbent party in close U.S. House races? Snyder (2005) interprets it as evidence of corrupt electoral manipulation, suggesting

[3] Urquiola and Verhoogen (2009) highlights similar problems with RD in the context of class-size caps, where strategic sorting by schools and students can invalidate the RD design.

Figure 1: Proportion of previous Democratic wins as function of Democratic vote margin, U.S. House 1946-2010



that the complexity of the process of collecting and tabulating votes in close elections leaves opportunities for incumbent candidates to somehow tamper with the results of close elections. Grimmer et al. (2012) expand on these ideas in analysis of a longer period of U.S. House races (1880-2008), showing that (particularly in the earlier period) candidates from the party that controlled local and state offices had a similarly substantial advantage; they suggest that part of the reason why "structurally advantaged candidates" disproportionately win close elections is that they are more successful in post-election legal battles.[4] While conceding that a convincing explanation for this sorting remains elusive, Caughey and Sekhon (2011) point to the ability of well-organized campaigns to obtain precise information about likely outcomes and to take extraordinary measures to secure victory in very close races.

We return to these explanations for sorting in U.S. House elections below. For now we note that the evidence of sorting in close U.S. House elections appears to cast doubt on the

[4]Grimmer et al. (2012) also suggest that structural advantages would be amplified in close elections because candidates exert unusually high effort in these situations, but it is unclear from the analysis how this would result in imbalances in RD analysis using standard procedures.

validity of RD as a strategy for estimating electoral effects not just in the U.S. House but in a much broader class of electoral contexts. As noted above, if certain types of candidates hold systematic advantages in close races, the key assumption of RD (quasi-random assignment to treatment in the neighborhood of the threshold) is violated. Furthermore, the explanations that have been offered so far would appear to apply quite generally to competitive elections in many settings, not just to elections for the postwar U.S. House. Although close U.S. House races in recent years attract more money and polling technology than we observe in most other electoral settings, there would seem to be at least as much scope for precise manipulation of outcomes in many other contexts. In legislative elections in many developing democracies, for example, electoral fraud is more common than in closely-monitored U.S. House contests (Lehoucq 2003; Simpser 2013). Polling technology is less widely used in most settings where researchers are interested in using RD to measure electoral effects, but in many of these settings the electorate is much smaller, such that candidates arguably have similarly precise information about likely outcomes. The existing evidence of systematic incumbent advantages in close U.S. House elections may therefore be thought to pose a quite general threat to the validity of RD-based electoral studies. In order to assess the nature of this threat we now assess whether the problematic patterns documented in the U.S. House are seen in other electoral settings.

III. Why focus on incumbency?

Before proceeding to our main analysis, we pause to justify our exclusive focus on (party) incumbency. While Snyder (2005) focuses on comparing the performance of incumbents and non-incumbents in close elections, other studies highlighting problems with RD in the U.S. House emphasize other comparisons: Grimmer et al. (2012) shows that winners of close House races disproportionately belonged to the locally-dominant party, especially in the period before World War II; Caughey and Sekhon (2011) test for imbalances in the largest set of background covariates, showing not just that districts where Democrats narrowly won the seat were more likely to have a Democratic incumbent but also that in those districts the Democrats received a higher vote share in previous elections, were more likely to be predicted to win, and spent more money (among other differences).[5] In order to evaluate the integrity of the RD design in a broader set of political settings, in principle we could test for anomalies in the full range of covariates examined in Grimmer et al. (2012) and Caughey and Sekhon (2011), including e.g. ideology of incumbent, share of spending by party, and party of the predicted victor. Instead, we focus on the party of the incumbent. The reason is that the other covariates for which previous studies have found anomalies are highly correlated with the party of the incumbent – so much so that after controlling for the party of the incumbent the evidence of imbalance in the other covariates disappears.

Table 1 makes the point with respect to the imbalances reported by Caughey and Sekhon (2011). In the leftmost column we report the full list of covariates for which Caughey and Sekhon (2011) find substantial imbalance. To document imbalance, they restrict attention to close elections (defined as those with a margin of less than 0.5%) and compute the mean difference for each covariate between districts in which the Democrat wins and districts where the Democrat loses; the middle column (labeled "Original Specification") reports the p-value corresponding to their test of the null hypothesis that this expected difference is zero.[6] In the rightmost column we report p-values from another analysis that differs only in that incumbency (i.e. "Democratic Win") is added as a control.[7] The fact that none of these p-values is below .1 indicates the high degree of collinearity between incumbency and each of these covariates. This suggests that focusing on incumbency may be sufficient: imbalance on incumbency produces imbalance on these other variables as well, and the purported imbalances on these other

[5]Caughey and Sekhon (2011) report that barely winners received slightly more campaign contributions and spent significantly more money than barely losers. However, their measure of expenditures includes post-election activity, raising the possibility that this imbalance could be explained by increased spending *after* an electoral victory – a phenomenon that would not pose a challenge to the assumptions of the RD design. We tested for this possibility by collecting contribution data from the Federal Election Commission separately for the periods before and after Election Day (expenditure data cannot be cleanly separated in this way). Focusing only on pre-election donations in close U.S. House races, we find little evidence of imbalance. However, as expected, barely winners do receive a significant bump in contributions *after* the election (results available upon request). These results suggest that the campaign finance imbalances reported by Caughey and Sekhon (2011) are largely driven by post election (i.e. post-treatment) activity.

[6]The p-values reported differ slightly from the ones depicted in Figure 2 of Caughey and Sekhon (2011) because we restrict attention to the subset of districts for which incumbency is defined. In general the reported p-values in the paper are slightly lower.

[7]We obtain the same results from a separate analysis where we regress each covariate on lagged incumbency, calculate the residuals, and test for balance on the residuals. This is nearly identical to our approach in the second column of Table 1 where we control for lagged incumbency.

Table 1: p-values from placebo tests in Caughey and Sekhon (2011) with and without controlling for incumbency

Dependent Variable	Original Specification	Including Dem Win $t-1$
Democratic Win $t-1$	.00	–
Democratic % Vote $t-1$	.10	.33
Democratic % Margin $t-1$	.03	.58
Incumbent D1 Nominate	.00	.60
Democratic Incumb in Race	.00	.58
Republican Incumb in Race	.00	.44
Democratic # Previous Terms	.08	.74
Republican # Previous Terms	.00	.10
Democratic Experience Adv	.00	.70
Republican Experience Adv	.00	.31
Partisan Swing	.00	.24
CQ Rating	.00	.47
Democratic Spending %	.01	.22
Democratic Donation %	.07	.53

NOTE: Cell entries are p-values for the variable *Democratic Win t* from linear regressions on the set of races in a 0.5% window, with robust standard errors. In the column labeled Original Specification the only regressor is *Democratic Win t*. In the column labeled Including *Democratic Win t−1* the two regressors are *Democratic Win t* and *Democratic Win t−1*.

variables go away once we account for incumbency. Put another way, even though Caughey and Sekhon (2011) report imbalances on many covariates, they have really only identified imbalance on one variable – an indicator for lagged incumbency – and the other imbalances come along for free because so many variables are correlated with lagged incumbency. For these reasons, we confine our subsequent analyses to lagged incumbency and lagged vote share.

IV. Do incumbents disproportionately win close elections?

We now turn to our main analysis. We analyze data for every partisan, single-winner, plurality/majoritarian electoral setting where data could be reasonably collected and assembled. This sample includes a broad range of politically significant national and local elections from 10 different countries. The data sets are listed in Table 2; in the Appendix we provide the source of each data set and details on how we handled issues such as redistricting and multi-party competition.[8] We follow Caughey and Sekhon (2011) in choosing a reference party for each setting (e.g. the Democrats in U.S. data sets; the Conservatives in the U.K. data sets) and calculating vote margins and incumbency status with respect to that party of interest. The vote margin for the reference party is the difference in vote share between the party of interest and the highest finisher among the other parties. Table 2 reports the number of races in each data set (as well as in the pooled data set) within 10%, 2%, and 1%. For example, a bandwidth of 1% includes all elections where the reference party won or lost by a margin of 1 point or less. In a case with only two parties, this would include all cases where the reference party won between 49.5 and 50.5% of the vote.

Table 3 presents analysis that aims to assess whether incumbent parties disproportionately win close elections in a variety of settings. Our basic strategy is to test for an "effect" of winning election at time t on incumbency status in time $t-1$. We carry out this placebo analysis using three common RD approaches. "Naive" analysis simply compares the mean values of the placebo outcome (an indicator for whether the reference party won the previous election) in narrow windows above and below the electoral threshold, testing for a difference

[8] In all settings we omit cases where the difference in vote share between the first and third place party is less than 5%; this is to avoid complexities emerging from close races involving more than two parties.

Table 2: Data analyzed

Setting	Obs. within 10%	2%	1%	Reference party
U.S., House of Reps, 1880-2010	5088	1085	568	Democratic
U.S., House of Reps, 1880-1944	3233	732	381	Democratic
U.S., House of Reps, 1946-2010	1855	353	187	Democratic
U.S., Statewide, 1946-2010	2202	498	250	Democratic
U.S., State Legislature, 1990-2010	5953	1204	582	Democratic
U.S., Mayors, 1947-2007	457	108	51	Democratic
Canada, Commons, 1867-2011	2553	576	278	Liberal
Canada, Commons, 1867-1911	759	205	102	Liberal
Canada, Commons, 1921-2011	1794	371	176	Liberal
U.K., Commons, 1918-2010	3414	675	336	Conservative
U.K., Local Councils, 1946-2010	10881	2123	1047	Conservative
Germany, Bundestag, 1953-2009	1260	262	131	CDU/CSU
Bavaria, Mayors, 1948-2009	928	195	87	CSU
France, Natl Assembly, 1958-2007	872	215	104	Socialist
France, Municipalities, 2008	458	104	59	Left
Australia, House of Reps, 1987-2007	349	73	39	Labor
New Zealand, Parliament, 1949-1987	330	57	27	National
India, Lower House, 1977-2004	1093	222	106	Congress
Brazil, Mayors, 2000-2008	1270	265	143	PMDB
Mexico, Mayors, 1970-2009	4016	801	404	PRI
All Races Pooled	41124	8463	4212	–

Note: See Appendix for details on each data set. The bandwidths are defined such that a bandwidth of 1% includes all elections where the reference party won or lost by a margin of 1 point or less.

in means.[9] "Local linear" analysis similarly tests for a jump in incumbency status at the threshold where the party of interest's vote margin changes from negative to positive, but it does so by fitting a regression interacting the vote margin with an indicator for whether the party of interest won the contest. "Polynomial" does the same with a third-order polynomial. For each type of analysis, we summarize the results by reporting the p-value on the test for a jump at the threshold, using italics to signal that the placebo treatment effect is negative, i.e. that incumbents appear to do *worse*. In the Appendix, we present these results graphically and for more specifications. Specifically, we present the results from local linear specification for all possible bandwidths between 0.5 and 5. These graphs present the point estimates for readers interested in interpreting the substantive size of the point estimates directly and also show that the results are robust across many specifications.

We begin by noting that our tests uncover convincing signs of imbalance in the U.S. House in the post-World War II period (row 3). The analysis in previous papers including Caughey and Sekhon (2011) has focused on "Naive" RD estimation, and we replicate their strong finding of imbalance not just for that specification but for the other RD specifications as well. We note, however, that for the U.S. House in the previous period as well as for the U.S. House in the entire period since 1880 we fail to find evidence of incumbent advantages in any specification at the .05 level.

Turning to the other U.S. contexts (statewide offices since 1946, state legislatures since 1990, and mayors since 1947), we find no evidence of an advantage for the incumbent party in any specification. This finding is particularly interesting given that existing explanations for incumbents' disproportionate success in the postwar U.S. House would seem to apply at least as strongly to these other contexts.

Turning to elections in other countries, we similarly fail to find any consistent evidence of an advantage to incumbent party candidates. Out of 96 tests shown for non-U.S. data, we do not once find a p-value below .05. When we pool all of the data into a single data set (bottom row of the table), we similarly find no evidence of incumbent advantages. The one case where the p-value is below .05 is the "Naive" analysis with a bandwidth of 1%, but given that the

[9] The naive analysis with a bandwidth of .5% is thus equivalent to a test for a difference in the binned means on either side of the threshold in Figure 1.

Table 3: Placebo tests: p-values for "effect" of party winning at time t on party winning at time $t-1$

	Naive		Local Linear			Polynomial	
Bandwidth =	*.5*	*1*	*1*	*2*	*5*	*5*	*10*
U.S., House of Reps, 1880-2010	.11	.06	.46	.29	.32	.29	.32
U.S., House of Reps, 1880-1944	*.70*	.96	*.59*	*.38*	*.93*	*.50*	*.65*
U.S., House of Reps, 1946-2010	.00	.00	.04	.00	.07	.00	.02
U.S., Statewide, 1946-2010	*.55*	*.79*	*.43*	*.38*	*.56*	*.50*	*.10*
U.S., State Legislature, 1990-2010	*.37*	.52	*.32*	.95	.59	*.78*	*.77*
U.S., Mayors, 1947-2007	–	.96	–	.81	.88	.37	*.62*
Canada, Commons, 1867-2011	*.29*	*.50*	*.32*	*.18*	*.09*	*.59*	*.17*
Canada, Commons, 1867-1911	*.59*	*.22*	.81	*.21*	*.19*	*.60*	*.18*
Canada, Commons, 1921-2011	*.30*	*.88*	*.18*	*.39*	*.17*	*.71*	*.35*
U.K., Commons, 1918-2010	.33	.09	.59	.61	.08	*.92*	.12
U.K., Local Councils, 1946-2010	.24	.06	.44	.27	.22	.17	.68
Germany, Bundestag, 1953-2009	.71	.54	.79	.48	*1.00*	.74	.84
Bavaria, Mayors, 1948-2009	*.13*	*.38*	*.21*	*.39*	*.16*	*.19*	*.30*
France, Natl Assembly, 1958-2007	.27	.79	.33	.55	.53	.47	.23
France, Municipalities, 2008	–	.31	–	.37	.14	.52	.24
Australia, House of Reps, 1987-2007	–	–	–	1.00	*.55*	.50	*.92*
New Zealand, Parliament, 1949-1987	–	–	–	–	.75	*.86*	.69
India, Lower House, 1977-2004	.49	.38	*.54*	*.98*	.20	*.97*	.86
Brazil, Mayors, 2000-2008	*.81*	*.81*	*.61*	*.58*	.78	*.64*	.97
Mexico, Mayors, 1970-2009	*.69*	*.96*	*.39*	*.68*	*.93*	*.93*	*.60*
All Races Pooled	.22	.02	.92	.59	.16	.46	.75

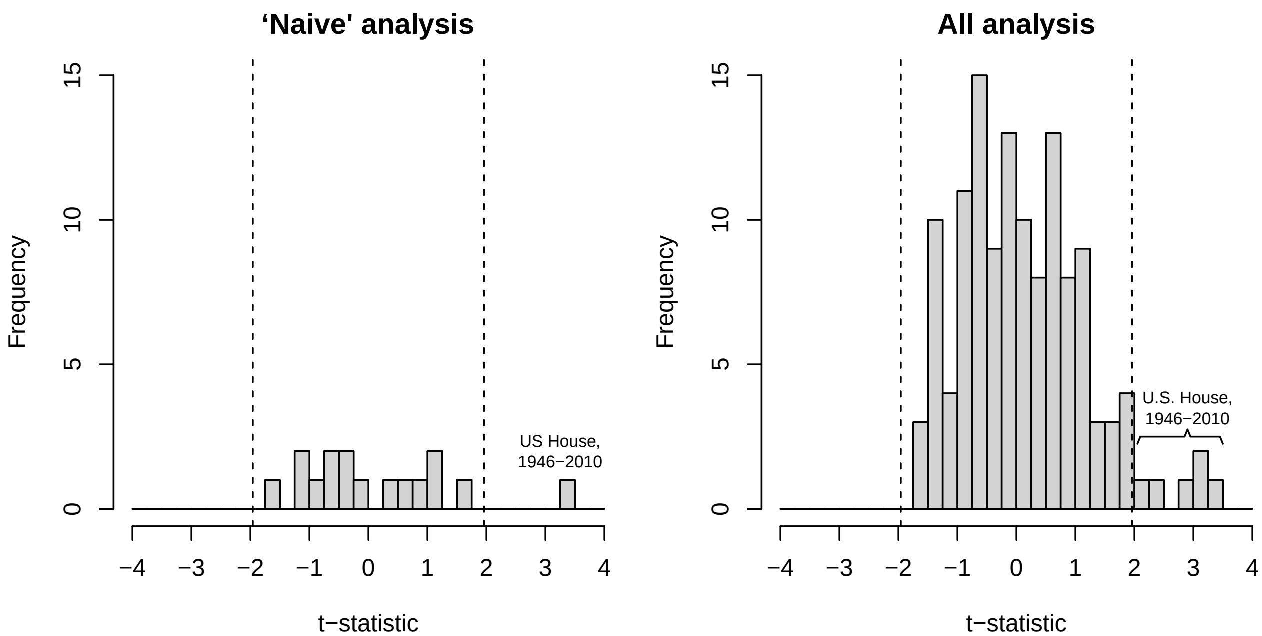
NOTE: Each entry gives the p-value of a two-tailed test of the hypothesis that the coefficient on *Treatment* is zero. Results not shown if there are insufficient datapoints within a given bandwidth, where the cutoffs are 40, 60, and 100 for naive, local linear, and polynomial. Robust standard errors used in all cases. Results in italics indicate that the point estimate is the opposite of what one would expect if incumbents disproportionately win close elections. Standard errors clustered by state-year for U.S. Statewide offices.

naive estimate is biased upwards we do not view this as convincing evidence of sorting.[10]

Figure 2 provides a graphical summary of the results in Table 3. In the left panel we plot the histogram of the t-statistics of the tests in the first column of Table 3 – "Naive" estimates of

[10]Put differently, because party performance is correlated over time a naive RD test should yield a significant result at any bandwidth given sufficient data, even if incumbents have no special advantages in close elections.

Figure 2: Summary of tests in Table 3



the difference in lagged victory rate between close winners and losers in a .5% bandwidth. The t-statistics are fairly evenly distributed around 0 except for a single outlier above 3: the U.S. House in the post-World War II period. In the right panel, we include all of the (non-pooled) tests from Table 3. Again the distribution appears to be roughly unimodal about 0, except for a right tail; *every one* of the t-statistics greater than 1.96 comes from the U.S. House in the post-World War II period.

As noted above, our placebo tests focus on (lagged) incumbency because our analysis in Table 1 suggests that incumbency accounts for most of the imbalances reported in existing studies for the U.S. House. It is good practice however to check for balance in the lagged running variable (Imbens and Lemieux 2008), i.e. the vote margin in the previous race. Table 4 reports results of the same tests using the same format as Table 3, where the outcome is the lagged vote margin rather than lagged incumbency status. The naive analysis shows imbalance in the U.S. House only at the 1% bandwidth for the post-World War II period; in no setting is there consistent evidence of disproportionate incumbent victories in close elections. Again, we present these results graphically and for many more specifications in the Appendix. Histograms of test statistics are displayed in Figure 3 and indicate a similar pattern to the one in Figure

2: t-statistics appear to be drawn from a unimodal density centered about 0.

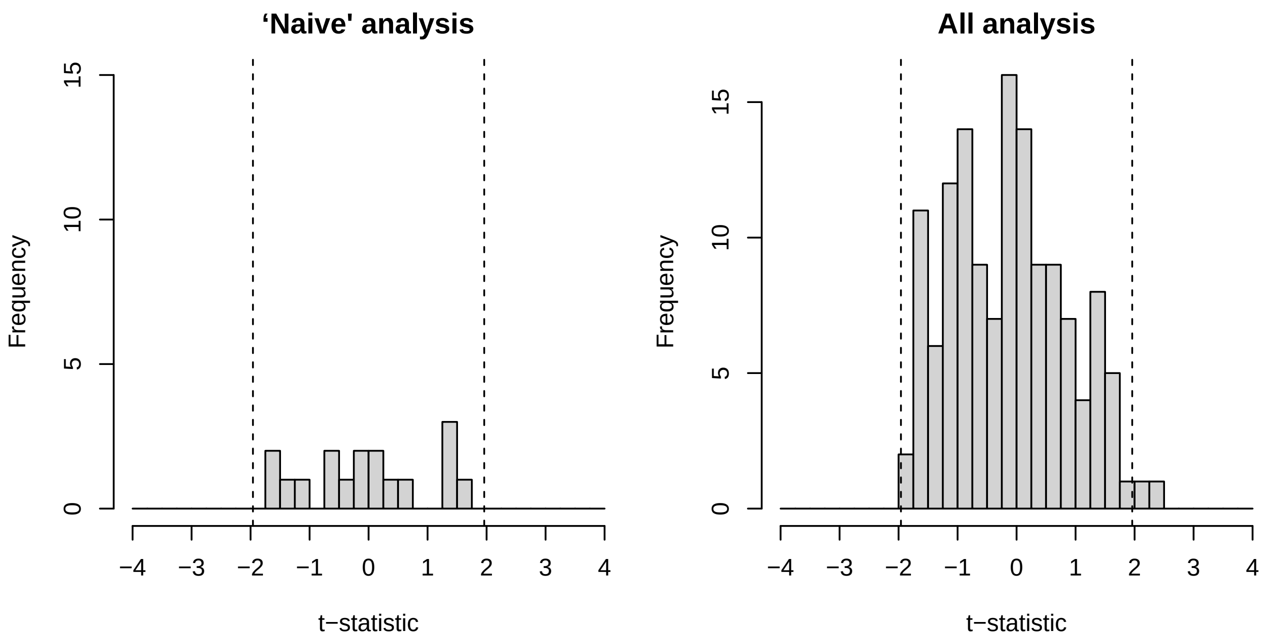
Table 4: Placebo tests: p-values for "effect" of party winning at time t on party vote margin at time $t-1$

	Naive		Local Linear			Polynomial	
Bandwidth =	*.5*	*1*	*1*	*2*	*5*	*5*	*10*
U.S., House of Reps, 1880-2010	.21	.15	.81	.51	.37	.77	.81
U.S., House of Reps, 1880-1944	.91	*.85*	*.77*	*.45*	*.95*	*.39*	*.57*
U.S., House of Reps, 1946-2010	.15	.04	.63	.16	.21	.29	.41
U.S., Statewide, 1946-2010	*.84*	.69	*.81*	*.82*	*.98*	.97	*.29*
U.S., State Legislature, 1990-2010	.75	.78	.92	*.91*	*.91*	*.89*	*.59*
U.S., Mayors, 1947-2007	–	*.11*	–	*.22*	*.42*	*.09*	*.10*
Canada, Commons, 1867-2011	*.12*	*.31*	*.13*	*.10*	*.06*	*.29*	*.08*
Canada, Commons, 1867-1911	*.26*	*.17*	*.38*	*.27*	*.08*	*.53*	*.12*
Canada, Commons, 1921-2011	*.21*	*.51*	*.20*	*.17*	*.17*	*.35*	*.19*
U.K., Commons, 1918-2010	.16	.11	.65	.43	.58	.67	.46
U.K., Local Councils, 1946-2010	.10	.02	.33	.12	.40	.08	.35
Germany, Bundestag, 1953-2009	.95	*.45*	.50	*.81*	*.29*	.98	*.37*
Bavaria, Mayors, 1948-2009	*.10*	*.39*	*.12*	*.30*	*.10*	*.23*	*.26*
France, Natl Assembly, 1958-2007	*.57*	*.39*	*.54*	*.26*	*.76*	*.34*	*.92*
France, Municipalities, 2008	–	.46	–	.83	.11	.92	.48
Australia, House of Reps, 1987-2007	–	–	–	.49	.30	.36	.18
New Zealand, Parliament, 1949-1987	–	–	–	–	.09	*.77*	.31
India, Lower House, 1977-2004	*.77*	.78	*.40*	*.78*	.21	*.88*	.89
Brazil, Mayors, 2000-2008	*.47*	*.77*	*.25*	*.33*	.52	*.32*	.95
Mexico, Mayors, 1970-2009	*.99*	*.77*	.83	.98	*.35*	.73	*.42*
All Races Pooled	.46	.25	*.95*	*.88*	.95	*.95*	*.50*

NOTE: See text for explanation of tests and notes to Table 4 for details on presentation.

In Table 5 we report the results of additional analysis based on the density test suggested by McCrary (2008). In these tests, we assess whether the density of incumbent-party candidates' vote share is smooth near the threshold; if candidates of the incumbent party disproportionately win close elections, we would expect a disproportionate number of incumbent-party candidates to narrowly win their contests. We first separate each data set according to whether the party of interest previously won the seat or not ("Incumbent" vs. "Non Incumb") and carry out the

Figure 3: Summary of tests in Table 4



McCrary test separately on each series, restricting attention to cases where the margin was within 10%. If incumbents disproportionately win close elections, we would expect a break in the density of the vote margin at 0 – a jump up for the sample of elections in which the party of interest held the seat and a drop down for the sample of elections in which the party of interest did not hold the seat. We do not generally find this pattern; even the results for the U.S. House in the post-World War II period are only borderline significant for the "Incumbent" series. We then recombine the two subsets while flipping the sign of the vote margin for the cases in which the party of interest was not the incumbent; for this combined data set, we would expect a bulge in the density where the adjusted margin is slightly above 0, indicating that the party of interest is likely to narrowly lose when it previously lost and likely to narrowly win when it previously won. As indicated by Table 5, we can reject the null of no density jump for all settings except the U.S. House after 1946.

V. Explaining the U.S. anomaly

Our analysis indicates that the U.S. House in the postwar period stands apart from other electoral settings: the incumbent party disproportionately wins close races in that setting but

Table 5: McCrary (2008) tests: p-values for null hypothesis of equal density on opposite sides of the threshold

	Incumbent	Non-Incumb	Pooled
U.S., House of Reps, 1880-2010	.80	.81	.93
U.S., House of Reps, 1880-1944	*.60*	*.60*	*.41*
U.S., House of Reps, 1946-2010	.07	.18	.05
U.S., Statewide, 1946-2010	*.43*	*.47*	*.26*
U.S., State Legislature, 1990-2010	*.83*	*.42*	*.41*
U.S., Mayors, 1947-2007	*.76*	.13	.39
Canada, Commons, 1867-2011	*.34*	*.62*	*.23*
Canada, Commons, 1867-1911	.65	*.14*	*.38*
Canada, Commons, 1921-2011	*.25*	.59	*.76*
U.K., Commons, 1918-2010	.44	.07	.10
U.K., Local Councils, 1946-2010	*.73*	.32	.46
Germany, Bundestag, 1953-2009	*.49*	.33	.64
Bavaria, Mayors, 1948-2009	*.26*	.83	.93
France, Natl Assembly, 1958-2007	*.62*	.03	.12
France, Municipalities, 2008	.	.91	.10
Australia, House of Reps, 1987-2007	*.72*	*.13*	*.13*
New Zealand, Parliament, 1949-1987	.40	1.00	.78
India, Lower House, 1977-2004	*.79*	.40	.58
Brazil, Mayors, 2000-2008	.45	*.37*	.83
Mexico, Mayors, 1970-2009	*.94*	*.63*	*.85*
All Races Pooled	*.81*	.42	.62

NOTE: See text for explanation of test and notes to Table 4 for details on presentation.

not earlier in the history of the U.S. House, not at the state and local level in the U.S., and not in a variety of national and local elections in other countries. This raises an important question: how should we interpret evidence of incumbent dominance in very close U.S. House elections, if not as an indication of a general feature of campaigns and elections?

We see two ways of interpreting evidence of sorting in the U.S. House. First, we could conclude that the imbalances we observe are the result of a systematic pro-incumbent bias, as suggested by Snyder (2005), Grimmer et al. (2012), and Caughey and Sekhon (2011); the fact that we find this bias only in the postwar U.S. House may point to the unusual amount of money and scrutiny devoted to these elections or other differences between these elections and others. Second, we could conclude that the imbalances we observe in very close U.S. House elections arose randomly: perhaps the determination of which candidates won and lost these elections *was* quasi-random, but by chance a disproportionate number of incumbent candidates ended up on the winning side. The question of which interpretation is most appropriate merits attention not just because of what it implies about the conduct of U.S. House elections but also because (along with the empirical analysis in the previous section) it sheds light on the question of how reliable RD estimates of electoral effects are likely to be. If the imbalance in the U.S. House were the result of systematic incumbent advantages in close elections rather than random chance, then (even despite the evidence presented in the previous section) we would be justifiably wary of the use of RD in electoral contexts with features similar to those of postwar U.S. House races. If on the other hand the apparent sorting in U.S. House elections is more likely a statistical "fluke," it should give us additional confidence in the validity of the RD assumptions in a broad class of electoral settings.

Based on our reading of existing explanations for sorting in the U.S. House and our understanding of U.S. congressional races, we argue for the latter view: imbalance in close U.S. House races seems more likely to be a statistical anomaly than the result of systematic advantages by incumbent candidates. Fundamentally, we arrive at our position by observing that each of the existing explanations for sorting in U.S. House elections seems highly doubtful when examined closely. Snyder (2005), Grimmer et al. (2012), and Caughey and Sekhon (2011) consider three related explanations for the disproportionate success of the incumbent party in close U.S. House races. One class of explanations has to do with what happens in a close election after the initial tally of votes. Incumbents may exert disproportionate influence on the recount process or may be better at winning the court cases that inevitably arise in very close races. Caughey and Sekhon (2011) rule out these explanations after finding that while recounts occur frequently in close races, they rarely reverse the initial result. This is consistent with the idea that incumbent-party candidates and challengers both bring substantial resources to election contests and thus incumbents cannot dominate at the recount stage.[11] Another class

[11]However, all of the 4 reversals identified and discussed by Caughey and Sekhon (2011) benefited the incumbent party, so recounts may explain some of the observed imbalance. If future work demonstrates that

of explanations has to do with outright fraud or electoral manipulation. Incumbent candidates may be more willing or more able than challengers to manipulate the outcome of elections through vote buying, ballot box stuffing, disfranchisement and exclusion of opponents, or other illegal means. If incumbents have extremely precise information about the expected election result, then fraud could potentially explain systematic incumbent party advantages in close elections. It is doubtful, however, that candidates could use fraud in the midst of a campaign in a sufficiently selective manner as to produce the type of sorting we observe in the postwar U.S. House, a point to which we return below in discussing conventional electioneering activities.[12] At any rate, as Caughey and Sekhon (2011) observe, fraud provides an unlikely explanation for sorting in the U.S. House because organized voter fraud in congressional elections is thought to be very rare in the recent period where we see imbalance (e.g., Lehoucq 2003).

A third class of explanations for imbalance in post-war U.S. House elections – preferred by Caughey and Sekhon (2011) – has to do with strategic, pre-election behavior by incumbent candidates and their campaigns. Incumbent candidates may possess extra resources relative to their challengers, allowing them to add an extra boost to their vote share if and only if they would otherwise expect to lose by a slim margin. At first glance, this story coincides with intuition. Incumbents keep close track of polling data, so they should know where they stack up relative to their opponent. They might have extra resources on hold that are costly to deploy (e.g. dipping into their own savings to launch more advertisements, calling in a one-time favor, making campaign promises they would prefer not to keep), so they will only use them

the imbalance in the House is primarily explained by recounts and court cases, there is a workable solution for applied researchers. If the initial vote tally is well-behaved but incumbents disproportionately prevail in recounts, then one can employ a "fuzzy" RD design in which the initial vote tally provides an instrument for the final election result. Note that this requires the usual fuzzy RD assumptions, including monotonicity and excludability (see for example Hahn, Todd and Van der Klaauw (2001)). The fuzzy RD also changes the estimand to the local average treatment effect for compliers, but in practice this estimand will be very close to the one from the sharp RD if recounts rarely reverse the initial vote result and therefore the rate of compliance is very high.

[12]The existence of fraud or electoral manipulation alone is not enough to violate the assumptions of the RD design, a point that may be relevant in electoral settings outside of the U.S. where fraud and manipulation are more common. Simpser (2013) demonstrates that even in democracies where fraud is widespread, its use does not appear to be strategic around electoral thresholds. For example, candidates often engage in fraud even when they are essentially assured of winning because extending their winning margin helps to show off their popularity, intimidate their opponents, and scare off future challengers. As a result, fraud itself does not constitute evidence that the RD assumptions are invalid (and similarly, evidence of covariate balance at the electoral threshold is not evidence that elections are free and fair).

if necessary. However, upon closer examination, the story becomes less plausible. To be clear, this class of explanations requires all of the following statements to be true:

1. Incumbents (or their allies) have access to some additional (but costly) campaign technology that they will only employ if necessary.

2. Incumbents (or their allies) have extremely precise information about the expected election outcome (we will clarify what we mean by *extremely precise*).

3. Challengers (and their allies) lag behind incumbents on at least one of these two dimensions (extra campaign resources or precise information about the expected result).

Statement 1 seems fairly plausible. Campaigns allocate resources and effort strategically: they make decisions about how much money to raise, how large of a staff to hire, how many speeches to give, how little to sleep, etc., based on their expectations about how close the election is likely to be (and thus how likely these efforts are to change the outcome).[13] Statement 3 is also plausible: even in close races, incumbents tend to have advantages in money and other resources; they might be expected to have a more extensive repertoire of campaign tactics to strategically deploy, as well as more precise information about when to deploy them. The fundamental problem with electioneering-based explanations of incumbent dominance in close elections is with statement 2: Even well-resourced campaigns do not possess anywhere near the precision of information about likely outcomes that is necessary for them to deploy campaign tactics in a way that would produce the observed sorting.

To see this, note that Caughey and Sekhon (2011) find imbalance in elections that were decided by a margin of 0.5 points or less, meaning that the Democratic share of the two-

[13]We note, however, that statement 1 becomes less plausible as Election Day draws nearer. In the final days and hours of the campaign, congressional candidates appear to exert as much effort as possible and utilize all available resources regardless of their standing in the race. There is rarely a stockpile of extra money and willing volunteers that will go unused unless the candidate decides to call upon them in the last hour to try to win a close race. Congressional candidates behave in ways that suggest that they are either extremely risk averse or that they care about their vote share *per se* over and above the binary electoral outcome. Just as corrupt candidates commit fraud even when they expect to win by a significant margin (Simpser 2013), U.S. House candidates may also exert significant campaign effort even when they expect to win in order to boost their ego, show off their popularity, and scare off future challengers. Some especially dramatic anecdotes relating to this behavior can be found in This American Life's 2012 episode entitled "Take the Money and Run for Office" available at `http://www.thisamericanlife.org/radio-archives/episode/461/take-the-money-and-run-for-office`.

party vote was between .4975 and .5025. Suppose incumbents (but not challengers) have access to a costly "secret weapon" that they can deploy to boost their vote share in close elections.[14] If strategic campaign activity explains the imbalance, there must be a significant share of incumbent candidates who will deploy their secret weapon if they expect to receive 49.9 percent of the vote, but will not bother doing so if they expect to receive 49.7 or 50.1 percent of the vote. It must be the case that, with an expected vote of 49.7%, the strategic incumbent would not bother, because she figures that she *cannot* win anyway; similarly, it must be that with an expected vote of 50.1%, the strategic incumbent can rest assured that she *will* win anyway and again will not bother deploying extra effort. This story, of course, assumes that incumbents can reasonably distinguish between situations where they expect to receive 49.7, 49.9, and 50.1 percent of the vote. The realities of political polling and campaigns cast serious doubt on the ability of candidates to obtain such precise expectations. Even expert statisticians utilizing all available information from polls and other sources on Election Day could not provide nearly enough precision for a candidate to be confident that she would win 50.1 as opposed to 49.9 percent of the vote.

Enos and Hersh (2013) provide evidence on the actual precision of campaign expectations by surveying Democratic candidates and campaign operatives in the run-up to the 2012 general election. On average, campaign workers mis-predict their vote share by 8 percentage points, and this unimpressive level of precision does not vary meaningfully across the status of the campaign worker (candidates and high-level managers are no better than volunteers and lower-level workers), the competitiveness of the race, the time until the election, or incumbent vs. challenger campaigns. For the 5 "toss-up" U.S. House races where Enos and Hersh (2013) surveyed the incumbent campaign, the operatives mis-predicted the election result by 10 percentage points, on average. Incumbent campaigns in close U.S. House races would have to predict their expected vote share within approximately one-quarter of one percentage point in order for strategic campaign behavior to explain the observed imbalance in House races, but actual candidates are about 40 times less precise than that.

Even if we do not trust campaign workers to reveal their true beliefs to researchers, we

[14] Or, equivalently, suppose that both incumbents and challengers have access to a "secret weapon" but it is more costly for challengers, such that only incumbents choose to use it.

can place an upper bound on the level of information that campaigns have by examining polls and statistical models in recent U.S. House elections. Klarner (2008) generates race-by-race predictions for the two-party vote share in every contested House election in 2008 using a plethora of polling data, election data from previous years, and other electoral and institutional factors known to influence congressional elections. These predictions were generated 100 days before Election Day, so they approximate the best possible predictions that a well-resourced candidate could have made at that point in time when they were making important, strategic, campaign decisions. On average, for contested races, these predictions miss the actual election result by 4.3 percentage points, and the average error exceeds 6 percentage points for the most competitive races. Closer to Election Day, campaigns can look at polls to obtain even more precise information about their expected vote share. The final poll or even the average of many late polls in a close U.S. House race in 2012, on average, missed the actual election result by about 2 percentage points.[15] Again, the stochastic and unpredictable nature of congressional elections means that campaigns, even those equipped with statistical models, have far less precision than they would need in order to create sorting. Taken together, this evidence casts serious doubt on the ability of strategic campaign activity to explain the observed imbalance in the post-war U.S. House.

Although campaigns can only predict their expected vote share within about 2 percentage points on the morning of Election Day, can they obtain more information during the course of the day and respond accordingly? Caughey and Sekhon (2011) write that U.S. House campaigns in close races "intensely monitor the vote as it comes in" and "react in real time" (pg. 400). However, the vote from a particular precinct is only reported after all polls have closed or after every registrant in the precinct has turned out (which only happens in tiny precincts), so campaigns could not possibly see the vote coming in and respond in time to alter the election result. The only information available to the campaigns during the course of Election Day comes from exit polls or from updated lists of which registrants have turned out in each precinct. Even with campaign workers stationed at every precinct and reporting the updated information to professional statisticians (something that is not routinely practiced in

[15]We conducted this analysis ourselves by collecting all of the polls available through Real Clear Politics.

congressional elections), campaigns could not possibly predict their expected vote share with the necessary level of precision to produce sorting. Surely, campaigns do respond to Election Day information in an attempt to improve their vote share. For example, they might deploy more cars in a part of the district where the weather is bad, or they might use the list of Election Day voters to determine which supporters to contact with a final get-out-the-vote message. However, the notion that campaigns keep significant resources in reserve, obtain extremely precise predictions of the election result, and call upon those reserves only when they otherwise expect to lose by a tiny margin appears implausible and inconsistent with the realities of modern congressional campaigns.

To summarize, existing explanations for the pattern of disproportionate incumbent-party success in close U.S. House elections are all unsatisfying: arguments based on post-election maneuvering, ballot box stuffing, or strategic deployment of campaign advantages are inconsistent with what we know about how election contests are actually fought. Given the difficulty of explaining the degree of incumbent success in close U.S. House elections, we favor the conclusion that the observed anomalies arose by chance. To be sure, if we look solely at close elections in the postwar U.S. House, we observe a degree of incumbent-party success that appears unlikely to have arisen randomly. However, given a large number of electoral settings it is *likely* that this degree of imbalance would emerge in one of them simply by chance. The analysis in this paper suggests that the postwar U.S. House may be that anomalous setting in which a high degree of imbalance arose by chance.

VI. Implications for the use of RD to estimate electoral effects

The fact that we fail to find problems in numerous electoral settings does not excuse researchers from defending the identifying assumptions of their empirical strategies and supporting them with evidence. When future researchers propose an RD design in a new setting, they should conduct the kinds of analyses we have conducted in this paper. At a minimum, they should clarify and justify their identifying assumptions, assess the robustness of their empirical results across specifications, and conduct tests for placebo effects of the treatment on the lagged outcome variable when possible. Additional placebo tests on the lagged running variable, lagged treatment variable, and other pre-treatment covariates along with tests for sorting based on McCrary (2008) would further bolster readers' confidence in the underlying assumptions and results (see also Imbens and Lemieux (2008) for a checklist of tests). The burden of proof is on the researcher to justify their assumptions and subject them to rigorous testing. A key advantage of the transparent RD design is that it lends itself to these numerous tests that follow directly from the identifying assumptions.

How should researchers proceed if they want to estimate electoral effects in the post-war U.S. House? We observe an incumbent-party advantage in extremely close elections in the post-war U.S. House but not in other electoral contexts. As we explain above, this suggests two possible interpretations of the U.S. House data. One is that there is something fundamentally problematic about the RD assumptions in U.S. House races, e.g. that despite evidence to the contrary there is in this context (and presumably not elsewhere) systematic pro-incumbent fraud in extremely close elections or that some subset of incumbents are able to obtain extraordinarily precise information about the expected outcome and use a "secret weapon" only when they would otherwise narrowly lose. The other interpretation is that the fundamental assumptions of RD are met in U.S. House races and the observed imbalance has happened by chance, just as in a randomized experiment where imbalances between treatment and control groups can arise by chance. We believe that the evidence, theory, and arguments presented in this paper support the latter interpretation. In this case, researchers hoping to estimate electoral effects in the modern U.S. House should proceed in a similar manner to researchers who discover chance imbalances in experimental data. While extraordinary care is required, valid inferences can still be made from imbalanced experiments.[16] One could adjust for imbalance by including lagged incumbency and other pre-treatment variables as covariates in the RD analysis or by pre-processing the data through matching or reweighting before conducting the RD analysis. Alternatively, researchers can also examine a "donut" RD design (Almond and Doyle 2011; Barreca et al. 2011), where they exclude the small sample of very close elections where imbalance exists. All of these fixes require additional assumptions, but the careful implementation of several approaches may produce valid inferences despite the presence of

[16]See Rubin (1973, 1979, 2009), Schochet (2010), and Miratrix, Sekhon and Yu (2013) for discussions of when and how valid inferences can be drawn from imbalanced experimental data.

imbalance. Even with the former interpretation where there is something fundamentally problematic about the RD assumptions in the U.S. House, the RD design may still be the best of all imperfect methods for estimating electoral effects in this important setting, and careful RD analysis may still produce better estimates than we could have otherwise obtained with other empirical strategies. As Caughey and Sekhon (2011) write, even in the case of estimating electoral effects in recent U.S. House elections, the RD design appears to be the best option: "Nevertheless, a comparison of the Lee RD estimator with traditional regression approaches to the incumbency advantage reveals that RD relies on weaker assumptions" (pg. 405).

VII. Conclusion

While our results should not induce complacency about the validity of RD designs in close elections, they should place the documented anomalies in U.S. House elections in the proper context. Our perception is that papers showing disproportionate incumbent successes in the U.S. House (particularly Caughey and Sekhon (2011)) have been highly influential among political scientists interested in estimating electoral effects. Absent careful analysis of other electoral contexts, one could easily conclude based on these papers that there is something fundamentally problematic about the use of RD to study electoral effects – particularly because there are many electoral settings where an incumbent might "steal" a close race more easily than in a highly scrutinized U.S. House election. Accordingly, in seminar discussions, conference panels, and referee reports one senses that documented anomalies in close U.S. House races have left the impression that RD designs are probably invalid in most electoral settings. For example, Einstein and Kogan (2012) (citing Caughey and Sekhon (2011) and Grimmer et al. (2012)) casts doubt on Gerber and Hopkins (2011)'s use of an RD design to estimate partisan effects on municipal spending in the U.S. by noting that RD findings rest "on the critical assumption that very close mayoral contests are indeed decided randomly, something that only rarely appears to be the case in practice" (pg. 9), despite no evidence that mayoral elections suffer from the same problem as U.S. House elections. Our experience is that these sentiments have become commonplace. Evidence of imbalance in the U.S. House has made some scholars suspicious of all RD-based studies, to the point where they lend more credence to other approaches, despite the fact that, as Caughey and Sekhon (2011) point out, the RD design "still makes weaker assumptions than the usual model-based alternatives" (pg. 406).

To our knowledge, this paper provides the most thorough and extensive assessment to date of the validity of the regression discontinuity design in electoral settings. Across more than 40,000 closely contested races in many different electoral settings, we find no systematic evidence of sorting around the electoral threshold. Conditional on being in a very close election, incumbents are no better at winning than challengers. We hope that these results will bolster confidence in estimates of electoral effects that arise from RD designs, so long as researchers exercise the appropriate levels of prudence and thoroughness. The theory and evidence assembled here suggests that the sorting and imbalance that has been discovered in the U.S. House is most likely a statistical fluke, the type of anomaly that is expected to arise by chance in a small fraction of cases. Recent concerns about the validity of electoral RD designs appear overblown, as we find no evidence that the underlying assumptions are categorically unsound.

References

Ade, Florian and Ronny Freier. 2011. "Divided Government Versus Incumbency Externality Effect: Quasi-Experimental Evidence on Multiple Voting Decisions." *DIW Berlin Discussion Paper No. 1121* .

Almond, Douglas and Joseph J Doyle. 2011. "After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays." *American Economic Journal: Economic Policy* 3(3):1–34.

Barreca, Alan I., Melanie Guldi, Jason M. Lindo and Glen R. Waddell. 2011. "Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification." *The Quarterly Journal of Economics* 126(4):2117–2123.

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 US House." *Electoral Studies* 28(1):123–128.

Caughey, Devin 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.

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

Dell, Melissa. 2012. "Trafficking Networks and the Mexican Drug War." *Working Paper* .

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.

Einstein, Katherine Levine and Vladimir Kogan. 2012. "Partisanship and Representation in Local Politics: A Re-evaluation and New Evidence from Mid-Size Cities.". Unpublished manuscript.

Enos, Ryan and Eitan Hersh. 2013. "Elite Perceptions of Electoral Closeness: Fear in the Face of Uncertainty or Overconfidence of True Believers." *MPSA Annual Meeting Paper* .

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.

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 Hirsh, Brian Feinstein and Daniel Carpenter. 2012. "Are Close Elections Random?" *Working Paper* .

Hahn, Jinyong, Petra Todd and Wilbert Van der Klaauw. 2001. "Identification and estimation of treatment effects with a regression-discontinuity design." *Econometrica* 69(1):201–209.

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.

Horiuchi, Yusaku and Andrew Leigh. 2009. "Estimating Incumbency Advantage: Evidence from Multiple Natural Experiments." *Working Paper* .

Imbens, Guido W. and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142(2):615–635.

Klarner, Carl. 2008. "Forecasting the 2008 US House, Senate and Presidential Elections at the District and State Level." *PS: Political Science & Politics* 41(04):723–728.

Lee, David S. 2008. "Randomized Experiments From Non-Random Selection in U.S. 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.

Lehoucq, Fabrice. 2003. "Electoral Fraud: Causes, Types, and Consequences." *Annual Review of Political Science* 6(1):233–256.

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.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142(2):698–714.

Miratrix, Luke W., Jasjeet S. Sekhon and Bin Yu. 2013. "Adjusting Treatment Effect Estimates by Post-Stratification in Randomized Experiments." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75(2):369–396.

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

Rubin, Donald B. 1973. "The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies." *Biometrics* 29(1):pp. 185–203.

Rubin, Donald B. 1979. "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies." *Journal of the American Statistical Association* 74(366a):318–328.

Rubin, Donald B. 2009. "Should Observational Studies Be Designed To Allow Lack Of Balance In Covariate Distributions Across Treatment Groups?" *Statistics in Medicine* 28(9):1420–1423.

Schochet, Peter Z. 2010. "Is Regression Adjustment Supported by the Neyman Model For Causal Inference?" *Journal of Statistical Planning and Inference* 140(1):246 – 259.

Simpser, Alberto. 2013. *Why Governments and Parties Manipulate Elections: Theory, Practice, and Implications*. Cambridge University Press.

Snyder, Jason. 2005. "Detecting Manipulation in U.S. House Elections." *Unpublished Manuscript* .

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.

Urquiola, Miguel and Eric Verhoogen. 2009. "Class-size caps, sorting, and the regression-discontinuity design." *The American Economic Review* 99(1):179–215.

Vogl, Tom. 2012. "Race and the Politics of Close Elections." *Working Paper Princeton University* .

Online Appendix: Graphs

Figure A1: Testing for imbalances in lagged incumbent victory. We exclude bandwidths that subset the data to fewer than 60 observations.

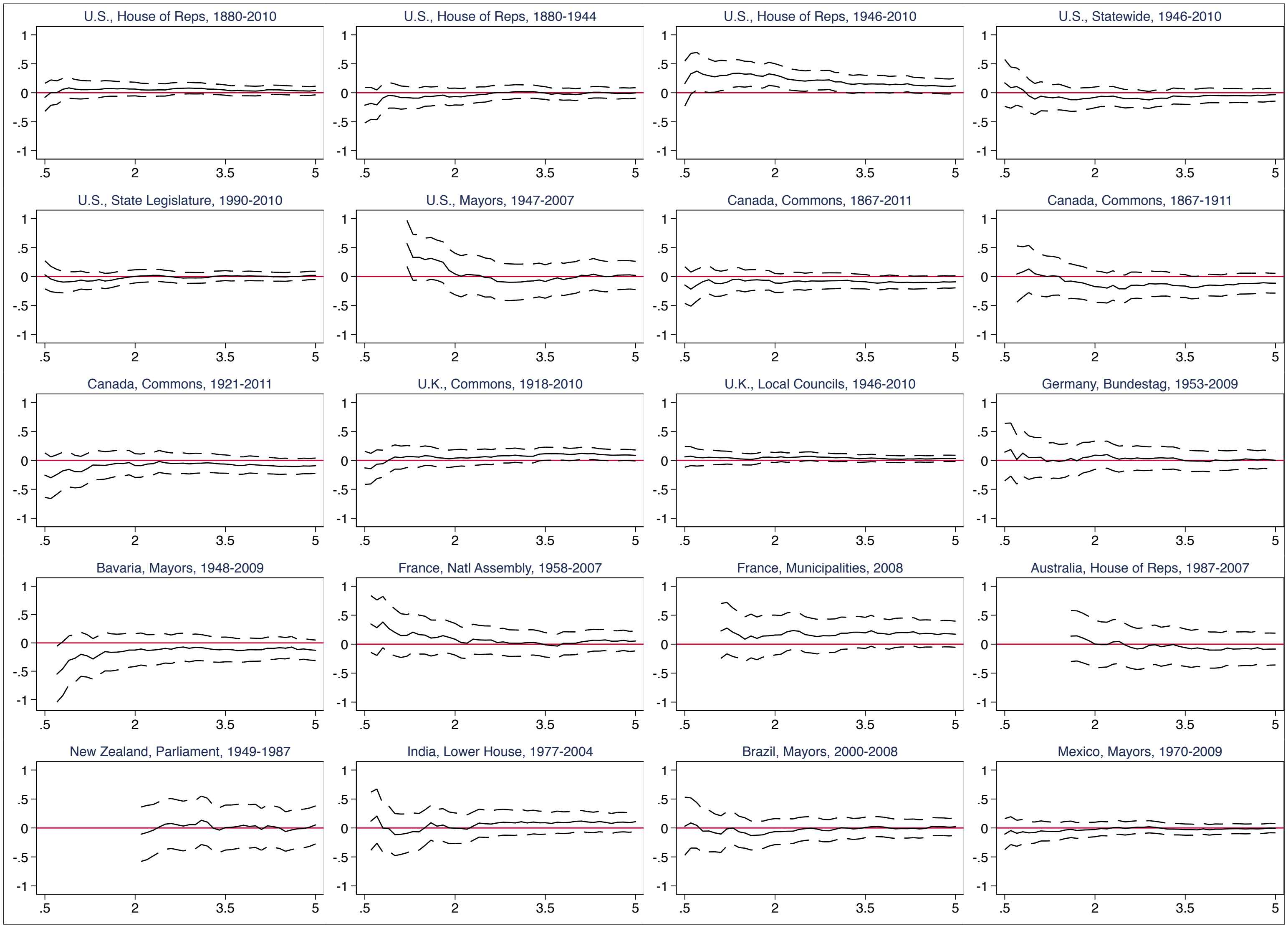


Figure A2: Testing for imbalances in lagged incumbent vote margin. We exclude bandwidths that subset the data to fewer than 60 observations.

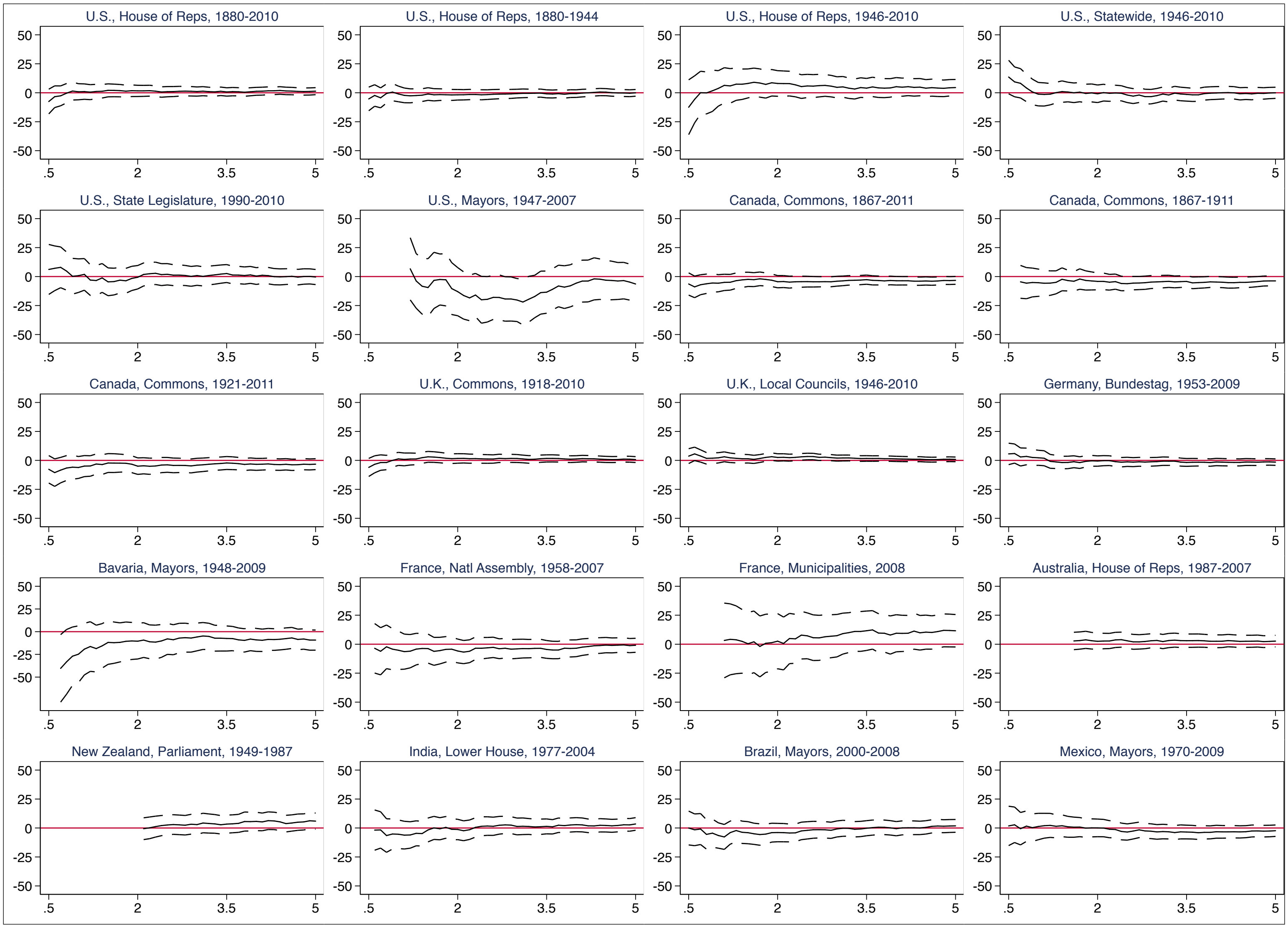


Figure A3: Testing for imbalances in lagged incumbent victory. All cases pooled.

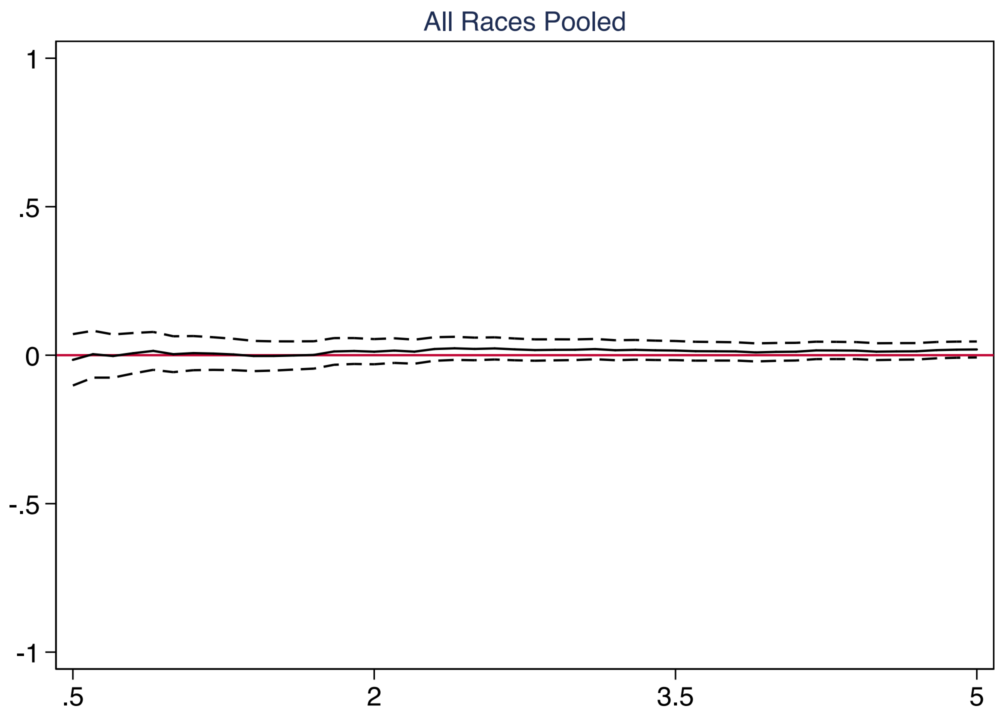
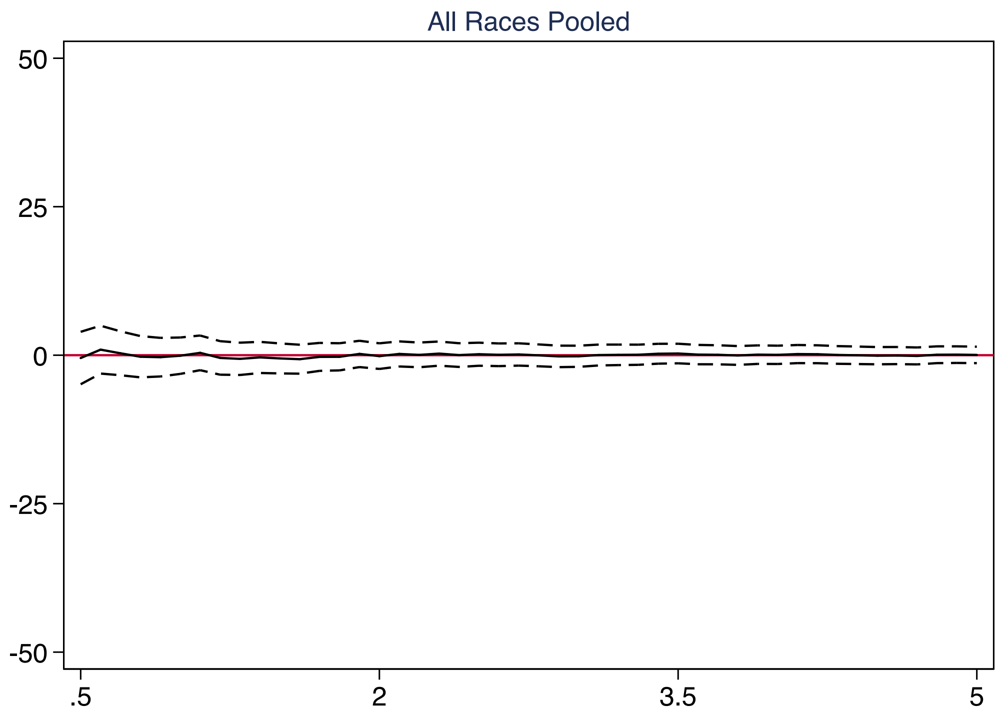


Figure A4: Testing for imbalances in lagged incumbent vote margin. All cases pooled.



Online Appendix: Data sources and definitions

A. U.S. State Legislative Elections

The U.S. State Legislative Election data comes from ICPSR Study 34297 (`http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/34297`). The data set provides election returns for all fifty states, 1967-2010. We exclude nonpartisan elections (most notably Nebraska's unicameral legislature) along with multi-member districts. We subset to outcomes from 1990-2010 in order to use only the most reliable information on off-cycle redistrictings. While state legislatures are nominally redistricted each decade in the year ending in '2', there have been a significant number of redistrictings in other years due to court cases and other extenuating circumstances. The data on redistricting from 1990 to present comes from Carl Klarner. This leaves us with 65,199 observations across 49 states.

B. U.S. Mayoral Elections

The U.S. Mayoral data was originally collected for Ferreira and Gyourko (2009) and has been extended by those authors in subsequent work. The extended data set contains mayoral election results for the years 1945-2007 in 834 cities, which includes non-partisan elections and elections in which members of the same party faced each other. We restrict to races where a Democrat faced a Republican, which leaves 2,396 observations spanning 494 cities.

C. Canadian House of Commons Elections

Data is provided by the Constituency-Level Elections Archive for elections to the House of Commons of Canada between 1867 and 2011.[17] The reference party is the Liberal Party of Canada. Members are elected in single member constituencies (ridings) by simple plurality. We exclude the few double-member ridings that existed in some provinces in the early periods. Redistricting is conduced by an independent commission every ten years. A riding is included in the analysis only when the riding boundary remains unchanged from the previous election. Data on historical boundary changes is provided by the Parliament of Canada, History of Federal Ridings Since 1867.

D. British House of Commons Elections

Data is provided by the Constituency-Level Elections Archive for elections to the British House of Commons between 1918 and 1997. Data for the elections in 2001, 2005, and 2010 are provided by the Electoral Commission and compiled by Rallings and Thrasher at the LGC Elections Centre at the University of Plymouth. The reference party is the Conservative Party. Members are elected in single member constituencies by simple plurality. We exclude the few multi-member constituencies that existed prior to 1950. Redistricting is conducted by a boundary commission every 8-12 years. A constituency is included in the analysis only when

[17]Ken Kollman, Allen Hicken, Daniele Caramani, and David Backer. Constituency-Level Elections Archive (CLEA; www.electiondataarchive.org), December 17, 2012 [dataset]. Ann Arbor, MI: University of Michigan, Center for Political Studies [producer and distributor].

the constituency name remains unchanged from the previous election; we cleaned and checked constituency names for consistency across elections. In the data we find that there are 1,850 unique constituencies across the 25 elections. Most of them experienced redistricting at some point during the sample period. The median constituency remains unchanged for seven elections.

E. British Local Elections

Data comes from the British Local Election Database published by Rallings, Thrasher, and Ware.[18] The reference party is the Conservatives. Analysis is based on single-member elections to county councils, district councils, and unitary authorities in England, Scotland, and Wales in the period 1945-2003. Wards are included in the analysis only when the ward boundary is the same as in the previous election.

F. German Bundestag Elections

Data is provided by the Federal Returning Officer (*Bundeswahlleiter*). The reference party is the Christian Democratic Union of Germany (CDU) together with its Bavarian sister party the Christian Social Union of Bavaria (CSU). Germany has a mixed electoral system where, since 1953, voters have two votes. The first vote is for a direct candidate for the constituency and the candidate who receives a simple plurality of first votes gets the direct mandate to serve in the Bundstag (SMD tier). Each constituency returns a single member. The second vote is for a party list and determines the proportion of seats a party receives in the Bundestag (PR tier). Analysis is based on the SMD tier races for the 12 elections between 1953 to 2009. Periodic redistricting is conducted by an independent election commission. A race is included in the analysis only when the constituency area remains unchanged from the previous election. Data on constituency areas is obtained from various years of the German election law (*Änderung des Bundeswahlgesetzes* 1949, 1964, 1972, 1976, 1979, 1985, 1989, 1990, 1993, 1996, 2001, 2005, 2008). Periodic redistricting often involves only a small subset of constituencies. 84 constituencies remain constant for all 12 elections. The median constituency remains unchanged for four elections.

G. Bavarian Mayoral Elections

Data has been collected, and provided to us, by Florian Ade and Ronny Freier and was originally used in Ade and Freier (2011). The data covers about 25 000 mayor elections in the state of Bavaria for the time period 1946-2009. A feature of these elections is important for the correct implementation of a correct analysis is the presence of a second (or run-off) ballot. If no candidate reaches the majority of 50% in the first round, a second round is held between the two leading candidates. If there is such a second round we use that in our analysis. We use the CSU as the reference party in our analysis. Also, we restrict the sample to contested elections with the top two candidates being from different parties. These restrictions leave us with a sample of a little bit less than 100 00 observations.

[18]Rallings, C.S., Thrasher, M.A.M. and Ware, L., British Local Election Database, 1889-2003 [computer file]. Colchester, Essex: UK Data Archive [distributor], June 2006. SN: 5319, `http://dx.doi.org/10.5255/UKDA-SN-5319-1`.

H. French National Assembly elections

Data is provided by CDSP (Centre de Données Socio-politiques) of Sciences Po and CNRS. The reference party is the Socialists. From 1958 to 1981 the results are aggregated by party label, meaning that the vote totals are incorrect in cases where multiple candidates from the same party compete. Analysis of the data from 1988 to 2007 indicates that this happened so rarely as to not pose a serious problem: two candidates of the same party label appeared in the second round in only about .6% of cases. (In the first round, which is rarely decisive, the rate was about 3.5%.) The election of 1986 was conducted via party-list proportional representation and was followed by a major redistricting; we thus omit the 1986 election and treat the periods before and after separately. (Other episodes of minor redistricting are dealt with by dropping observations in which the lagged outcomes took place under different boundaries.)

Legislative elections in France take place in two-round contests: if no candidate wins a majority of votes in the first round, then a second round is held in which all candidates receiving less than 12.5% of the first-round vote are eliminated. (The threshold was 5% before 1977.) We define the running variable in these cases based on the decisive round – the round in which the winner was declared.

I. French municipal elections

Data is provided by the Ministry of the Interior. Analysis is based on the 2008 election in cities with at least 3,500 inhabitants. The electoral system in this setting is not single-member plurality as it is in the other settings we study: municipal elections in France take place between lists of candidates rather than between individual candidates, and the electoral system is nominally proportional rather than plurality rule. Including these elections in the analysis makes sense, however, because the electoral system confers a large "winner's bonus" of 50% of the seats to the winning list (the remainder of seats are distributed proportionally among all of the lists), such that the winner of a close contest between two lists ends up with a large majority and can thus choose the mayor. If sorting is a problem in SMP elections, therefore, one would expect to find it here as well.

Due to the large number of parties and inconsistent labeling of parties across years, we use as the reference party the "Left", meaning lists labeled by the Ministry of the Interior in 2008 as Socialist, Communist, "miscellaneous Left", extreme Left, Green, or union of the Left; in 2001, the corresponding labels are Left, "miscellaneous Left", extreme Left, and Green.

As in legislative elections in France, municipal elections take place in two rounds. (At the municipal level, lists winning less than 10% of the vote are eliminated.) We take the same approach, basing the running variable on the decisive round.

J. Australian House of Representatives Elections

Data on Australian House of Representatives Elections from 1987 to 2007 is from the Australian Electoral Commission as assembled and cleaned by Horiuchi and Leigh (2009). The reference party is the Australian Labor Party. Australia has essentially a two-party system with the Labor Party on the left and several other parties typically forming a coalition on the right. Voting is by a preferential system (or instant runoff) where voters rank candidates, allowing for the calculation of a two-party preferred vote for the top two candidates. Our analysis focuses on the Labor Party's share of the two-party preferred vote.

Redistricting in Australia is conducted by an independent commission before every election, but the changes are typically small. Between the 1990 and 2010 elections (when redistricting data is available) 59 percent of districts were not changed at all before an upcoming elections, only 26 percent of districts were changed by 10% or more (meaning that 10% of the voters in that election were new to the district), 16 percent of districts were changed by 20% or more, 10 percent of districts were changed 30% or more, 6 percent of districts were changed by 40% or more, and only 3 percent of districts were changes by 50% or more. We cannot restrict our analysis based on the extent of redistricting in a particular electoral division or year, because the placebo outcomes may have potentially influenced the redistricting process. However, given the minimal extent of redistricting in each election, attenuation resulting from redistricting is likely to be minimal.

K. New Zealand House of Representatives

Data is provided by the Constituency-Level Elections Archive for elections to the New Zealand House of Representatives between 1946 and 1987.[19] The reference party is the New Zealand National Party. Members are elected in single member districts by simple plurality. Redistricting is conduced by an independent commission every fifth year. A district remains in the analysis only if its name has not changed from the previous election, which we use to approximate large redistricting events.

L. Indian Lower House Elections

Data is provided by the Election Commission of India for elections to the lower house of parliament (Lok Sabha) between 1977 and 2004. The reference party is the Indian National Congress (INC). Candidates are directly elected in single member constituencies by simple plurality. Constituency boundaries remain unchanged during this period (apart from a few changes in the state boundaries).

M. Brazilian Mayoral Elections

Data is provided by the Supreme Electoral Tribunal (*Tribunal Superior Eleitoral*) for mayoral elections in 2000, 2004, and 2008. The reference party is the Brazilian Democratic Movement Party (*Partido do Movimento Democrático Brasileiro*). Mayors are elected by simple plurality in each municipality. The vast majority of municipalities only have one round, but large municipalities can have a run-off election and for those municipalities we use the results from the second round. There is no redistricting during this period. In a very small number of cases the municipality names change and these cases are excluded (following cleaning to identify unique names across election years).

[19]Ken Kollman, Allen Hicken, Daniele Caramani, and David Backer. Constituency-Level Elections Archive (CLEA; www.electiondataarchive.org), December 17, 2012 [dataset]. Ann Arbor, MI: University of Michigan, Center for Political Studies [producer and distributor].

N. Mexican Municipal Elections

State-by-state municipal election data for Mexico was collected by Melissa Dell for Dell (2012) among other studies. The original data "are from Mexico Electoral-Banamex and electoral results published by the Electoral Tribunals of each state. For 11 states, data on the total number of eligible voters, required to calculate turnout, are not reported" (Dell 2012: 34). Elections are multi-party; we use PRI as the party of interest.