Political Analysis (2011) 19:385–408
doi:10.1093/pan/mpr032

Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942–2008

Devin Caughey
Travers Department of Political Science, University of California, Berkeley, CA 94720-2370

Jasjeet S. Sekhon
Travers Department of Political Science, Department of Statistics and Director, Center for Causal Inference and Program Evaluation, Institute of Governmental Studies, University of California, Berkeley, CA 94720-2370
e-mail: sekhon@berkeley.edu (corresponding author)

Following David Lee's pioneering work, numerous scholars have applied the regression discontinuity (RD) design to popular elections. Contrary to the assumptions of RD, however, we show that bare winners and bare losers in U.S. House elections (1942–2008) differ markedly on pretreatment covariates. Bare winners possess large *ex ante* financial, experience, and incumbency advantages over their opponents and are usually the candidates predicted to win by *Congressional Quarterly*'s pre-election ratings. Covariate imbalance actually worsens in the closest House elections. National partisan tides help explain these patterns. Previous works have missed this imbalance because they rely excessively on model-based extrapolation. We present evidence suggesting that sorting in close House elections is due mainly to activities on or before Election Day rather than postelection recounts or other manipulation. The sorting is so strong that it is impossible to achieve covariate balance between matched treated and control observations, making covariate adjustment a dubious enterprise. Although RD is problematic for postwar House elections, this example does highlight the design's advantages over alternatives: RD's assumptions are clear and weaker than model-based alternatives, and their implications are empirically testable.

1 Introduction

In recent years, social scientists have come to recognize the centrality of research design to causal inference. As a consequence of this "design-based" revolution, scholars have increasingly turned to quasi-experimental designs that take advantage of arbitrary or haphazard manipulations of causal variables (Dunning 2008; Robinson, McNulty, and Krasno 2009). The regression discontinuity (RD) design, which exploits situations in which units are assigned to treatment on the basis of an arbitrary cutoff score, has experienced a particular surge in popularity. The RD design offers the promise of clean quasi-experimental answers to knotty inferential questions. Indeed, it is one of the few observational designs that has been shown to be able to recover experimental benchmarks (Cook, Shadish, and Wong 2008; Green et al. 2009; Shadish et al. 2011). This promise is exemplified in Lee's (2008) pioneering use of RD to estimate the incumbency advantage in U.S. House elections, a work that has received over 300 citations and

Authors' note: An appendix and supplementary materials for this article are available on the *Political Analysis* Web site. We thank David Lee for generously providing replication files for Lee (2008). We are grateful to Scott Adler, David Brady, Gary Jacobson, Keith Poole, and Jonathan Wand for sharing their data with us. We thank the editors, the anonymous reviewers, Henry Brady, Andy Eggers, Andrew Gelman, Don Green, Jens Hainmueller, Luke Keele, Winston Lin, Walter Mebane, Jr., Eric Schickler, Laura Stoker, Dan Tokaji, Rocío Titiunik, and Rob Van Houweling for providing helpful comments, and Peter Ryan for helping to shape the project in its formative stages. Willa Caughey, Mona Fang, Julia Gettle, and Sarah Weiner provided excellent research assistance.

been widely emulated in political science (e.g., Butler 2009; Eggers and Hainmueller 2009; Gerber and Hopkins 2011).[1]

As we demonstrate, however, RD is no panacea for the selection and other biases that plague observational estimates of the incumbency advantage. Using an original data set on U.S. House elections and candidates, we show that pretreatment characteristics exhibit a discontinuous "jump" at the threshold separating Democratic victories from Democratic defeats. In fact, the outcomes of very close elections can be predicted with a high degree of accuracy based on such *ex ante* indicators as the partisanship of the previous incumbent, the financial resources of the candidates, and *Congressional Quarterly*'s pre-election race ratings. Covariate imbalance around the cut-point casts doubt on RD's crucial identifying assumption that treatment assignment changes discontinuously at the threshold but the potential outcomes do not.

Why are some candidates able to eke out narrow victories? Even in competitive elections, U.S. House candidates are not evenly matched. Partisan tides may make the out-party candidate more competitive than usual, but our data show that the incumbent party's candidate nearly always has more political experience and more money. These observable factors are likely correlated with other unobserved advantages, such as party organization, political skill, or the preferences of constituents. In the closest elections, candidates have every incentive to make maximal use of their resources, and not coincidentally, almost three-quarters of razor-close elections break towards the party that already holds the seat. This bias towards the incumbent party is even evident in open-seat elections, though open seats are too infrequent to support firm conclusions. Because Democrat-held seats tend to be in jeopardy when there is a national swing towards the GOP (and vice versa for Republican-held seats), close victories for each party occur overwhelmingly in good years for the opposite party.

Why do Lee (2008) and other works applying RD to U.S. House elections not detect the sorting at the cut-point? The primary reason is that the models they use to test for covariate continuity extrapolate using data far from the threshold and are therefore insufficiently sensitive to the unique dynamics of close elections. Contrary to the expectation that covariate distributions should converge in the limit, we find that in the immediate neighborhood of the cut-point elections actually appear to become *less* consistent with random assignment—that is, selection becomes more severe. The divergence right around the cut-point fits with the well-known fact that candidates, voters, and other relevant actors behave very differently in highly competitive elections. To determine whether the pro-incumbent bias is due to postelection manipulation, we collected detailed data on a sample of close elections. We find that the first reported vote total winner is almost never reversed in subsequent counts, suggesting that the bias is the result of activities on or before Election Day rather than recounts or fraud committed after the election.

This paper outlines these findings and explores their methodological and substantive implications, using Lee's (2008) paper on the incumbent party advantage in the U.S. House as a reference point. We find that the pro-incumbent bias in close House elections is so severe that the true incumbent party advantage could easily be a fraction of the size of the RD estimate and possibly nonexistent. It is not our intention, however, to denigrate the RD design, which we believe to be a powerful inferential tool. RD's identifying assumptions are weaker *a priori* than those underlying model-based alternatives, and unlike other observational methods, they have clear observable implications that can be tested. Even when the assumptions are not satisfied, as in the case of U.S. House elections, the RD design makes weaker assumptions than the usual regression methods that rely on all elections.

We begin with a discussion of the theory and assumptions behind the RD design and a survey of its applications to elections. Section 3 demonstrates that covariate imbalance at the cut-point is too extreme for RD's key identifying assumption, the "smoothness" of the potential outcomes, to be plausible in the case of post-war U.S. House elections. Section 4 examines potential explanations for the sorting we discover and discusses their substantive political implications. The penultimate section considers what we can learn from this imperfect RD and compares it to the Gelman and King (1990) incumbency advantage estimator. We conclude with a summary of our findings and recommendations for researchers.

[1]The insight that the RD design could be applied to elections seems to have first appeared around 2001, in separate working papers by Lee (2001) and Pettersson-Lidbom (2001).

2 RD Designs: Theory, Application, and Interpretation

The RD design, though developed by the psychologist Donald Campbell in the early 1960s (Thistlethwaite and Campbell 1960; Campbell and Stanley 1963), received scant attention from statisticians and social science methodologists until about 15 years ago (Cook 2008). Since then, appreciation of the useful statistical properties of the RD design and of its applicability in a wide variety of contexts has grown to the point where it is now an important "arrow in the quasi-experimental quiver" (Angrist and Pischke 2010, 13). An RD design is potentially applicable in any situation where the relationship between an observed continuous variable (the "forcing," "assignment," or "running" variable) and the causal variable of interest (the "treatment" variable) exhibits a discontinuous jump at a certain threshold (the "cut-point"). The original use of the RD design, for example, was to estimate the effect on student achievement of winning a scholarship awarded to those who scored above a certain threshold on a standardized test (Thistlethwaite and Campbell 1960). The intuition behind this approach is that if student achievement jumps discontinuously at the cut-point—that is, if bare winners of the scholarship differ markedly from bare losers—then the difference can be attributed to the effect of the scholarship.

2.1 *Theory and Assumptions of the RD Design*

Hahn, Todd, and Van der Klaauw (2001) and Lee (2008) formalize the RD design in the language of the Neyman-Rubin model (Splawa-Neyman, Dabrowska, and Speed 1923/1990; Rubin 1974). In this framework, the causal effect of treatment $T \in \{0, 1\}$ on unit i is $\tau_i = Y_i(1) - Y_i(0)$, where $Y_i(1)$ denotes the potential outcome of i under treatment and $Y_i(0)$ the potential outcome under control. If the potential outcomes are distributed smoothly at the cut-point, the RD design estimates the average causal effect of treatment at the cut-point, $Z_i = c$:

$$\tau_{RD} \equiv \mathbb{E}[Y_i(1) - Y_i(0)|Z_i = c] = \lim_{Z_i \downarrow c} \mathbb{E}[Y_i(1)|Z_i = c] - \lim_{Z_i \uparrow c} \mathbb{E}[Y_i(0)|Z_i = c]. \tag{1}$$

Given the importance of the smoothness assumption, it is worth discussing in detail. Why, for example, do we find this assumption to be plausible in the case of a scholarship awarded to those who surpass a certain threshold on a standardized test? The answer is often couched in terms of the "manipulability" of the forcing variable, but as Lee and Lemieux (2010, 293) note, the key question is how *precisely* the score can be manipulated. In the test example, test-takers are certainly "manipulating" their scores, by trying to answer as many questions correctly as they can or, perhaps, to achieve the minimum score needed to win the scholarship. It is even possible that some may be cheating by, for example, glancing at their neighbor's answers. But unless some test-takers whose scores are near the threshold know with absolute certainty whether each of their answers is right or wrong, there is at least some element of randomness in their final score. Lee (2008, 684) argues that as long as the random component is "non-trivial," there is a randomized experiment hidden in the neighborhood of the cut-point, at least asymptotically.

Given finite data, the crucial issue is whether the random component is "non-trivial" relative to the precision with which the relationship between Z and Y can be estimated. Consider the example of roll call votes in the U.S. House of Representatives, discussed in McCrary (2008). In this case, the forcing variable is the percent of members voting in favor of a bill, and treatment is the passage of the bill (note that the forcing variable varies discretely in increments of about 0.2%). For a number of reasons—for example, majority-party agenda control (Cox and McCubbins 2005) or Riker's (1962) "size principle"—it is reasonable to expect bills that barely pass to be more common than those that barely fail. And, in fact, McCrary (2008) demonstrates that the empirical density of bills is discontinuous at the threshold of passage, dropping precipitously just below 50% but jumping up again just above. The conclusion to be drawn in this case is not that there is no random component to how much support a bill receives. Rather, the random element is not large enough, and our measure of bills' support not fine grained enough, for bills on either side of the cut-point to be plausible counterfactuals for one another.

One of RD's primary advantages over other quasi-experimental approaches, such as instrumental variables, is that the implications of its identifying assumptions are more directly observable. Two basic kinds of tests of the validity of the RD design have been proposed: continuity of pretreatment covariates and irrelevance of covariates to the treatment–outcome relationship (e.g., Lee and Lemieux 2010, 296–7).[2] The former approach is much more closely tied to the notion that the RD design takes advantage of an implicit natural experiment in the neighborhood of the cut-point. While the continuity of the potential outcomes cannot be tested directly, an implication of random treatment assignment is that the treated and control groups should not differ systematically in their pretreatment characteristics.

Perhaps the most straightforward such test is the density test suggested by McCrary (2008), which is analogous to checking whether the ratio of treated to control units in an experiment departs significantly from chance. In addition, one can also test for systematic relationships between treatment and covariates, which should be nonexistent under randomization. Randomization checks are more complicated to implement in the RD design than in experiments because the former implies covariate balance only in the limit—that is, at the cut-point itself. If the relationships (independent of treatment) between covariates and the forcing variable are weak or nonexistent, then covariates should be balanced on either side of the cut-point. If the relationships are strong, however, covariate imbalance may be present even when the smoothness assumption holds because balance is only expected in the limit. Thus, unlike true randomized experiments, the analysis of RD designs often requires modeling assumptions not justified by the presumed randomization itself (Green et al. 2009).

The second major approach to testing the robustness of the RD design is to examine whether conditioning on baseline covariates alters the estimated treatment effect (for implementations of this approach, see Lee 2008; Pettersson-Lidbom 2008; and Gerber and Hopkins 2011). Under random assignment, there should be no systematic association between covariates and treatment, so conditioning on them should not (asymptotically) change the estimates.

2.2 *The RD Design Applied to Elections*

In recent years, more than 20 works (see Table 1) have used the RD design to estimate the effect of election outcomes, and even more have appeared since we began work on this paper.[3] Four of the works focus on U.S. House elections, but national and subnational legislatures in a variety of countries are also represented, as are local contests and even unionization elections. A number of these studies focus on the incumbent party advantage, but a variety of other political, economic, and social outcomes have been examined, including legislators' voting patterns, the geographical distribution of government appropriations, the results of elections for other offices, legislators' personal wealth, and the political success of female candidates.

Nearly all of the published papers present at least some empirical evidence that the assumptions of RD are satisfied in their application of the design. The most common sort of evidence provided is a test of the continuity of the distribution of covariates at the cut-point. The vast majority of these continuity tests entail regressing a covariate on a third- or fourth-order polynomial in vote share or margin, interacted with dummies for treatment status (sometimes temporal or geographic fixed effects are included as well). A statistically insignificant coefficient for the treatment dummy is taken as evidence in favor of local random assignment. Rarely is the order of the polynomial explicitly justified, though Lee, Moretti, and Butler (2004) or Lee (2008) are often cited in support of a quartic specification.

Many of the works also provide graphical evidence in the form of plots of covariates' local means, which is helpful in visually assessing the fit of the model. Two of the papers estimate covariate discontinuities using local nonparametric regression, as suggested by Hahn, Todd, and Van der Klaauw (2001) and Imbens and Lemieux (2008). A couple of the papers reverse the left and right sides of the model, regressing either treatment status or the outcome variable on one or more covariates while controlling for vote margin, with the idea that the relationship should not be significant. Cellini, Ferreira, and Rothstein (2010) plot a coarse histogram of vote share, a rough approximation of the density test advocated by

[2] Additional specification checks have been suggested. For example, Pettersson-Lidbom (2008, 1051–2) tests whether pretreatment variables are significantly associated with his outcome variable, controlling for a polynomial in vote share, under the logic that "the pretreatment characteristics should not have any effect at the discontinuity because the pretreatment characteristics should [be balanced] close to the threshold." Another common test is to look for discontinuities elsewhere in the range of the forcing variable.

[3] In addition, a number of works have cited Lee (2008) as an exemplary application of the RD design, and many of them use the data set from the paper to illustrate the design or demonstrate particular techniques (e.g., Imbens and Lemieux 2008; McCrary 2008; Angrist and Pischke 2009; Imbens and Kalyanaraman 2009; and Lee and Lemieux, 2010). By contrast, Snyder (2005) in an unpublished working paper provides empirical evidence that bare winners and losers are not comparable. He examines congressional elections in which an incumbent is running for reelection.

McCrary (2008). A few of the papers, including Hainmueller and Kern (2008) and Titiunik (2009), test for covariate differences in moments higher than the mean, as one should in a true randomized experiment. In addition, many of the papers estimate the treatment effect conditional on the covariates, as a means of both improving the precision of the estimates and evaluating their robustness. This robustness check is generally employed in conjunction with, rather than as a substitute for, direct tests of covariate continuity.

In summary, electoral applications of RD have become common in political science as well as economics. Nearly all these works present at least some empirical evidence for the randomness of the elections they examine. On the other hand, many of the papers evince the attitude that unless there is widespread fraud or other distortions to the vote count, elections should be random almost by definition. The following quotation summarizes the emerging consensus: "there is no reason to expect the winners and losers of elections decided by razor-thin margins to systematically differ in any way" (Eggers and Hainmueller 2009, 523).[4] Unlike Lee (2008), however, few of these studies collect data on and test for the continuity of lagged measures of the treatment variable (usually the party that currently controls the office or body). This may be a serious omission, for as we shall see, this covariate is likely to be particularly diagnostic of nonrandom sorting around the cut-point. Often even less justification is given for estimation technique and functional form. The problem of model dependence is hardly unique to RD and is generally more serious in conventional regression-based observational studies. But given the importance of accurately modeling the data right around the cut-point, model specification can be very consequential for estimation of both treatment effects and covariate discontinuities.

3 Sorting in Close Elections

As Lee (2008, 684) shows, elections will be randomly decided in the limit as long as the vote share has "a non-trivial random chance component." There is nothing inherent about elections, however, that dictates that this model must hold in practice. Listokin (2008) and McCrary (2008), for example, provide evidence that vote share is subject to precise manipulation in corporate shareholder elections and congressional roll call votes, respectively. Even if the model holds asymptotically, in any actual observed sample of elections, the random chance component may well be too "trivial" relative to the degree of sorting for the RD design to work in practice. In this section, we demonstrate that the winners and losers of razor-close U.S. House elections differ markedly on observed characteristics, strongly suggesting that their unobservable potential outcomes differ as well. Because the ability of U.S. House candidates to sort at the cut-point dominates the random component of the vote margin, the distribution of potential outcomes is most likely discontinuous as well.

In order to evaluate the comparability of bare winners and losers in U.S. House elections, we constructed a unique data set that substantially improves and expands upon the data used by Lee (2008) and others who have applied RD to U.S. House elections. Lee's data, which cover the years 1946–1998, are derived primarily from "Candidate and Constituency Statistics of Elections in the United States, 1788–1990" (ICPSR 1995). Although Lee corrected many of the numerous errors in the original data, a great deal of miscoded or missing data remained in the data set he used in his analysis. The data imputation and name-matching methods Lee used resulted in unreliable measurement of key variables, especially pretreatment covariates such as the candidates' previous terms in office. In addition, Lee treats a district as having been redrawn only in years ending in "2," when in fact redistricting often does not coincide with decennial reapportionment. Even if the measurement error on pretreatment covariates is purely random, it has the effect of concealing discontinuities at the threshold.

Our data set corrects these errors based on authoritative sources and adds dozens of demographic, political, and candidate variables to the data set, allowing for a richer and more accurate portrait of U.S. House elections. We provide a detailed description of how our data set was constructed in online Appendix A. Particular attention was paid to collecting complete and accurate data on close elections. Unless otherwise noted, the analyses that follow are based on elections between 1942 and 2008 in districts whose boundaries were not redrawn since the previous election or before the next election (i.e., elections for which

[4]Eggers and Hainmueller (2009) present compelling evidence for the validity of this statement in the British elections they examine.

Table 1 Review of applications of the RD design to elections

Author (year)	*Type of work*	*Elections*	*Statistical evidence for validity*
Albouy (2009)	NBER working paper	U.S. House and Senate	No explicit evidence provided
Broockman (2009)	*Political Analysis*	U.S House	Regresses outcome in window on two key covariates
Brollo and Nannicini (2010)	IGIER working paper	Brazilian mayoral races	(1) Histogram of vote margin (2) Density test of vote margin (3) Census covariates regressed on cubic in margin
Butler and Butler (2006)	*Political Analysis*	U.S. Senate	(1) Covariates regressed on quartic in margin (2) Mean differences in windows of 10%, 5%, and 2% (no incumbent party dummy)
Cellini, Ferreira, and Rothstein (2010)	*Quarterly Journal of Economics*	U.S. school bond referenda	(1) Coarse histogram of vote share (2) Covariates regressed on cubic in share (3) Change in covariates regressed on cubic in share
DiNardo and Lee (2004)	*Quarterly Journal of Economics*	NLRB races	(1) Local means of covariates plotted (2) Covariates regressed on quartic in margin
Eggers and Hainmueller (2009)	*American Political Science Review*	U.K. House of Commons	Covariate continuity tested with local linear regression (no incumbent party dummy)
Ferreira and Gyourko (2009)	*Quarterly Journal of Economics*	U.S. mayors	Covariates regressed on cubic in margin; refers to working paper and Lee (2008) for more evidence
Hainmueller and Kern (2008)	*Electoral Studies*	German Bundestag	(1) Covariates regressed on cubic in margin (2) Covariates included as controls in outcome model (variance of party share but no incumbency variable)
Hays and Franzese (2007)	Conference paper	OECD parliaments	No evidence presented, but paper is in progress
Gerber and Hopkins (2011)	*American Journal of Political Science*	U.S. mayoral races	Using cubic in margin, each covariate is: (1) included as control in outcome model (2) regressed on margin (no incumbent party dummy)
Horiuchi and Leigh (2009)	Working paper	Australian lower house	No evidence presented, but paper in progress
Jacob and Singhal (2010)	Working paper	Indian Parliament	Tests for covariate discontinuity using cubic and local linear (Imbens–Kalyanaraman) regression

Continued

Table 1 Continued

Lee (2008)	*Journal of Econometrics*	U.S. House	(1) Plots local means of covariates (2) Regresses covariates on quartic in margin (3) Covariates included in outcome model
Lee, Moretti, and Butler (2004)	*Quarterly Journal of Economics*	U.S. House	Regresses covariates on quartic in margin (no incumbent party dummy)
Leigh (2008)	*European Journal of Political Economy*	U.S. governors	No evidence presented
Pettersson-Lidbom (2008)	*Journal of the European Economic Association*	Swedish local governments	Using 1st- to 4th-order polynomials in margin: (1) covariates included in outcome model (2) party win regressed on covariates and margin
Titiunik (2009)	Working paper	Brazilian mayors	(1) *t* and KS tests in 5% and 3% windows (2) covariates include in outcome model
Trounstine (2011)	*Legislative Studies Quarterly*	U.S. city councils	Compares political experience of winners and losers in 5% window
Uppal (2009)	*Public Choice*	Indian state legislatures	(1) Difference of means in 5% window (2) covariates regressed on quartic in margin
Uppal (2010)	*Economics and Politics*	U.S. state legislatures	(1) Local means of covariates plotted (2) covariates regressed on quartic in margin

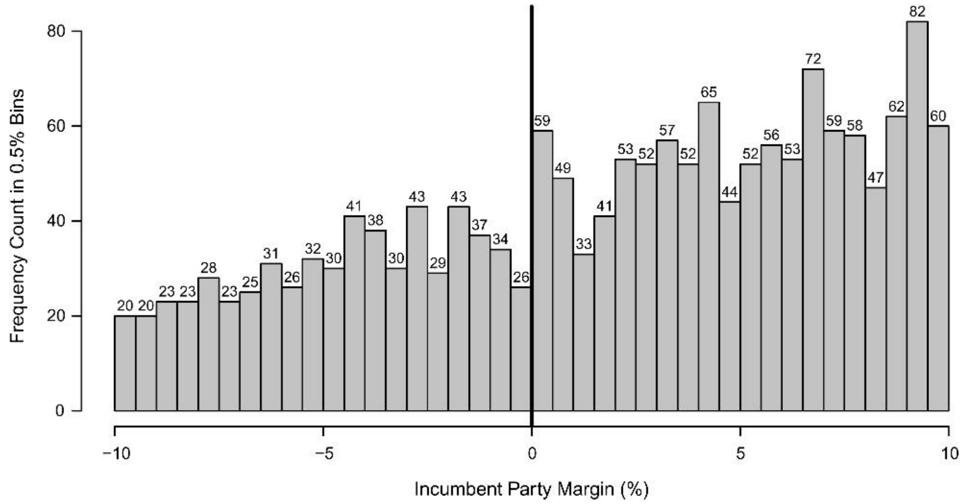


Fig. 1 Histogram of the incumbent party's margin in U.S. House elections (bin width = 0.5%).

both lag and lead data are available).[5] Except for the extended time span and the corrections made to the redistricting variable, this is equivalent to the sample of elections analyzed in Lee (2008).

3.1 *Density of the Forcing Variable*

The pattern of sorting is revealed most starkly with a histogram of the incumbent party's margin in U.S. House elections. As Fig. 1 shows, almost three-quarters of races decided by less than half a percentage point are won by the party that also won the previous election in that district. The difference between the 0.5% bins adjacent to the cut-point is highly significant and much larger than the difference at any other point.[6] The discontinuity is also evident if a local linear smoother is fit to the bin counts. Using 0.5% bins highlights the window around the cut-point in which balance worsens most dramatically, but binning in intervals of 0.25% or 1% yields the same basic results.

McCrary (2008) conducts a very similar test of sorting of U.S. House elections, with one crucial difference: pooling all seats together, he examines whether the density of the *Democratic* margin changes discontinuously at the cut-point. In results not reported here, we confirm his finding that it does not. This indicates that neither party has a general advantage in eking out close elections. Figure 1, however, shows that testing the density of the Democratic margin misses an important pattern: each party possesses such an advantage *in the seats it already occupies*.[7] Democrats are much more likely to barely win seats they won last time than to barely lose them, and Republicans have the same advantage in their seats. Since the partisan bias in the density are roughly equal, they cancel out when the forcing variable is defined as Democratic margin.

3.2 *Covariate Imbalance*

We now examine whether bare winners and losers in U.S. House elections are equivalent in terms of key pretreatment covariates. As one would expect given the pattern in Fig. 1, we find that winners and losers are not comparable, especially with respect to covariates correlated with the party of the current incumbent.

[5] We collected data for 1944 only on very close races and for 1942 only in districts where a close race occurred in 1944. Including the 1944 races, the date range for which lag and lead data are available is 1944–2006. Elections in which the Democratic margin was undefined (e.g., because the top two candidates were both Democrats) were excluded from the sample.

[6] The t statistic at the actual threshold is above 4. The test was replicated for simulated cut-points between −99.5% and 99.5% and the largest t statistic was about 3.

[7] See Fig. 1 in online Appendix C for histograms of *Democratic Margin t* broken down by incumbent party. Following Lee (2008), we define *Democratic Margin* as the difference between the main Democratic candidate's vote total and that of her nearest opponent, as a percentage of all votes cast.

Lee (2008, 690) compares Democratic winners and losers on six pretreatment variables.[8] He shows that differences between the two groups of candidates diminish as the margin shrinks (Table 1) but do not entirely disappear within a margin of 5%, the smallest window he reports. Lee argues that this is "to be expected" since "the sample average in a narrow neighborhood [is] a biased estimate," requiring extrapolation to the threshold (688). Using a fourth-order polynomial in Democratic margin (with a logit link as appropriate) fitted separately to each side, Lee estimates the value of the covariate distributions at the cut-point itself, showing graphically that the fitted lines converge almost perfectly (686–9, Figs. 2–5). He also finds no "effect" of Democratic victory in election t on Democratic victory or vote share in election $t-1$ (690–1). Based on this evidence, Lee concludes that in very close elections, the victorious party is "as good as randomized" (Lee 2008, 691, footnote 16).

Figure 1 suggests, however, that covariate balance dramatically worsens right around the cut-point, a pattern that even a flexible polynomial fitted to the entire range of the data is unlikely to model accurately. Table 2 demonstrates this point in greater detail. Even in a 0.25% window, in which the median election was decided by fewer than 200 votes, the bias is still large and statistically significant. If elections were truly random in the limit, then we would expect the percentages in the rightmost column to approach equality as the margin shrinks. Instead, they diverge, suggesting that races decided by less than half a percentage point are actually *less* consistent with random assignment than elections decided by slightly larger margins. Thus, a simple treated–control comparison in a small window may actually understate the difference in the limit.

Figure 2 provides summary of covariate balance in U.S. House elections decided by less than half a percentage point (i.e., less than 50.25–49.75% in a two-candidate race).[9] The results are substantially similar if discontinuities are estimated using local linear regression at a wide range of bandwidths, including those selected by the Imbens–Kalyanaraman algorithm (Fuji, Imbens, and Kalyanaraman 2009). For example, using this algorithm, the z-scores for the discontinuities in both campaign spending and donations are over six.[10] One clear pattern that emerges from the balance figure is that the most imbalanced covariates are those specific to the race in question, particularly those closely related to the party of the current incumbent. By contrast, measures of the partisanship of state officials (governor, secretary of state) and of stable district characteristics (normal presidential vote, demographic variables) are at most modestly related to the party of the winner, at least in the 0.5% window.[11]

Far from being randomly decided, the outcomes of very close elections are actually quite predictable. For example, of the 44 very close races in which the *Congressional Quarterly (CQ)* October race predictions favored one party, *CQ* correctly called the outcome in 31 races (70%).[12] As Table 3 indicates, the *CQ* ratings for election t even do a decent job of predicting the outcome in election $t+1$. Even if we examine only the 24 very close races in which *CQ* favored neither party, on the rationale that these races are the pure tossups, we find that the incumbent party pulled out a victory in nearly two-thirds of cases. In fact, the outcomes of close *CQ* tossups are actually more strongly associated with the outcome of the *previous* election than of the subsequent one.

Perhaps an even better measure of the expected outcome of an election than expert ratings is the money raised by the two candidates. Indeed, under some models of campaign contributions, the share of money raised by each candidate corresponds perfectly to their probability of winning the election (e.g., Snyder 1990). Money also proxies for the general balance of resources between the two candidates. We collected data on the campaign spending of the two candidates (available since 1972), as well as on the donations received by each candidate before Election Day (available since 1980); the results are similar using either measure.

[8]These variables are as follows: Democratic vote share $t-1$, Democratic win $t-1$, previous House terms of the Democratic candidate, previous House terms of the Democrat's main opponent, the Democrat's previous House races, and the opponent's previous House races.

[9]Balance tests were also conducted on a number of other census variables, the results for which mirror those of the demographic variables included in the table. No substantive differences arise if the reported p values in the figure are adjusted for multiple comparisons using the Benjamini and Hochberg (1995) method for false discovery rate (FDR) correction. The FDR correction changes the results little because most of the variables are highly correlated with incumbency.

[10]We use the boundary-optimal triangular (edge) kernel for local linear estimation (Cheng, Fan, and Marron 1997). See online Appendix E for further details.

[11]Our finding that governor and secretary of state are not significantly imbalanced is at odds with the findings of Grimmer et al. (2011), though these authors examine a longer time period than we do.

[12]*CQ* race predictions are available for most years since 1954.

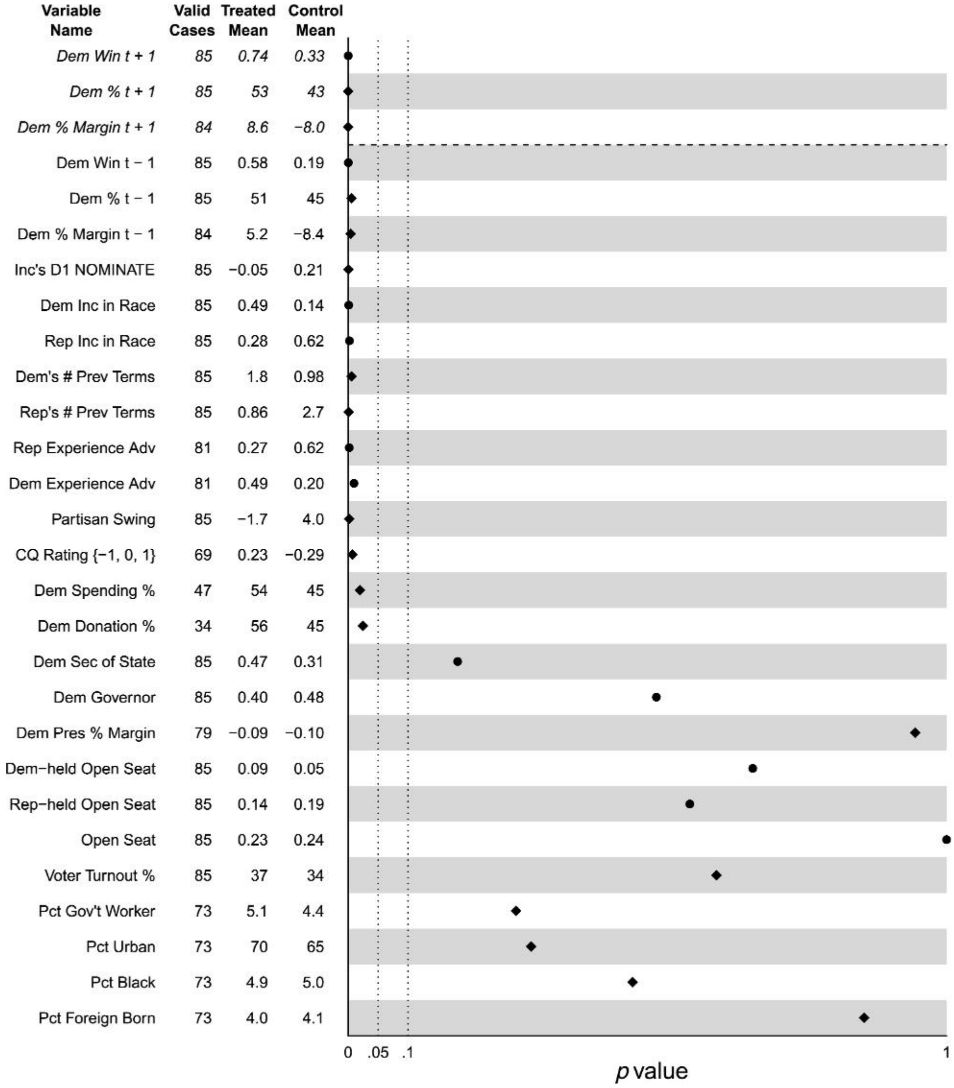


Fig. 2 Covariate balance between treated (Democratic win: $n = 43$) and control (Democratic loss: $n = 42$) in a 0.5% window. The first three variables listed are posttreatment outcome variables. The p values for dichotomous variables (circles) are from Fisher's exact test. Exact Wilcoxon rank sum tests were used for continuous and ordinal variables (diamonds). All p values are two-sided. Calculations are based on all cases with non-missing values for the variable.

Both the incumbent party candidate and the candidate who ultimately wins the election—who are usually the same person—generally enjoy a substantial financial advantage over their opponent. Even in *CQ* tossup races decided by less than 0.5%, two-thirds of winning candidates in such races spent more money than their opponent did. The same pattern holds if we examine all tossup races (regardless of the election outcome) or all races in the 0.5% window (regardless of how *CQ* classified them). In sum, financial resources are not equal in close elections, and the candidate with more money usually ends up winning.

In addition to their advantages in terms of material resources, the winners of close elections also tend to have more political experience than the candidates they defeat. In almost 80% of close races, one candidate had a substantial advantage in terms of previous political experience.[13] The more experienced candidate

[13]We code a candidate as having a substantial experience advantage if the candidate (a) is the incumbent representative and her opponent is not a former member of Congress or (b) has held another elected office but her opponent has not. In 10% of close elections, both candidates had previously served in Congress.

Table 2 Cross-tabulation of current and lagged Democratic victory, 1942–2008

	Dem Loss $t-1$	*Dem Win* $t-1$	*% Dem-held*
	Margin $< 2\%$ ($N = 320$)		
Dem loss t	107	65	38
Dem win t	74	74	50
	Margin $< 1\%$ ($N = 167$)		
Dem loss t	62	23	27
Dem win t	37	45	55
	Margin $< 0.5\%$ ($N = 85$)		
Dem loss t	34	8	19
Dem win t	18	25	58
	Margin $< 0.25\%$ ($N = 45$)		
Dem loss t	14	3	18
Dem win t	13	15	54

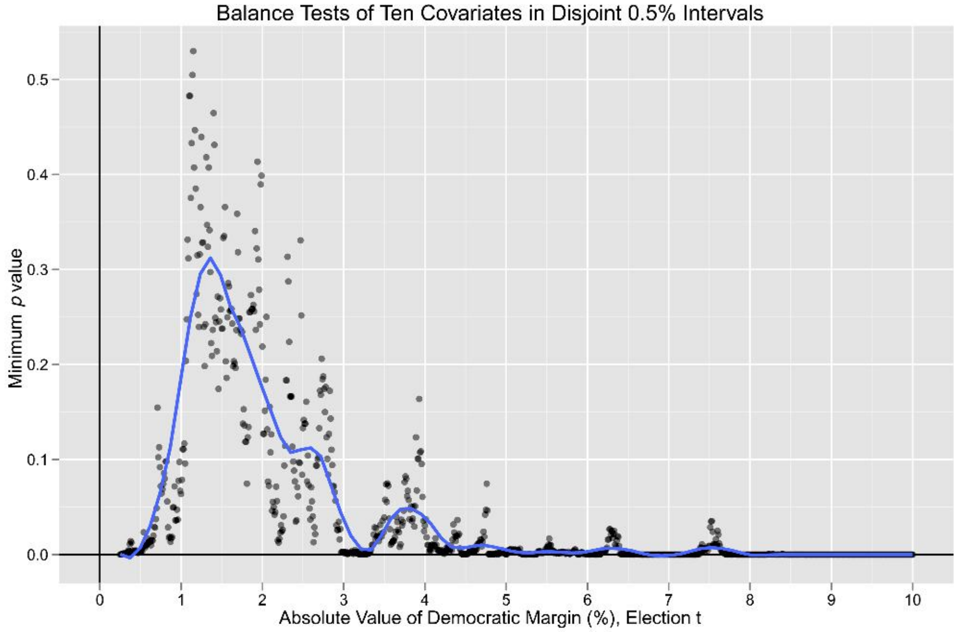
won 70% of these races. Incumbents running for reelection did even better, winning 72% of the time (see Krasno, 1994 on the poor quality of House challengers relative to the Senate). Open seats constitute only a quarter of races in a 0.5% window, and the small sample size ($n = 20$) limits our ability to draw firm conclusions about them. There is little affirmative evidence, however, that bare winners and losers are equivalent in open seats either. Despite the small sample, candidate spending is significantly imbalanced in the 15 open-seat races for which data are available. The incumbent party won 12 of 20 open seats in the 0.5% window. Although the confidence interval for this proportion (0.6) includes 0.5, it is also statistically indistinguishable from the proportion of incumbent party victories in nonopen seats. It is possible that sorting is less prevalent in open seats, but the data cannot support a firm conclusion either way.

As Schickler, Pearson, and Feinstein (2010, 682, footnote 25) observe, the kind of House districts in play in a given election differs markedly depending on the partisan tide in that year. For example, in 1958, a year with a strong pro-Democratic national tide, all six seats decided by less than 0.5% had been won by a Republican in 1956. The opposite held in 1994, a strong Republican year, when all five closely decided elections occurred in Democrat-held seats. When combined with our finding that the incumbent party usually pulls out close elections, this pattern has a further implication: Republican bare victories tend to occur in years with a pro-Democratic swing, and Democratic bare victories predominate in pro-Republican years. This could easily cause upward bias in the RD estimate of the incumbent party advantage due to regression-to-the-mean effects. On the other hand, we also find that about 20% of incumbents who barely win reelection retire in the next election (possibly because they face higher-quality challengers after a narrow reelection; Krasno 1994, 162), whereas no victorious challengers in our sample retire. By depriving the winning party of the personal advantages of incumbency, the higher rate of retirement by incumbents who barely survive may cause downward bias in the RD party incumbency estimate.

One potential counter to the foregoing analyses is that covariate discontinuities should be modeled rather than estimated nonparametrically in a small window around the cut-point. After all, the RD design only predicts convergence in the limit, not in any given window. On these grounds, Lee models trends in the data with a fourth-order polynomial and finds no statistically significant discontinuities. A major problem with this argument, however, is that covariate distributions in the immediate neighborhood of the cut-point are diverging rather than converging. The divergence of covariate distributions is illustrated in Fig. 3, which was constructed by calculating covariate imbalance in a moving 0.5%-wide window

Table 3 Cross-tabulation of *CQ rating* and *Democratic victory* in elections decided by less than 0.5%.

	Dem win $t-1$ %	*Dem win* t %	*Dem win* $t+1$ %
CQ: Republican favored ($n = 23$)	17	30	48
CQ: neither favored ($n = 25$)	24	52	48
CQ: Democrat favored ($n = 21$)	90	71	71



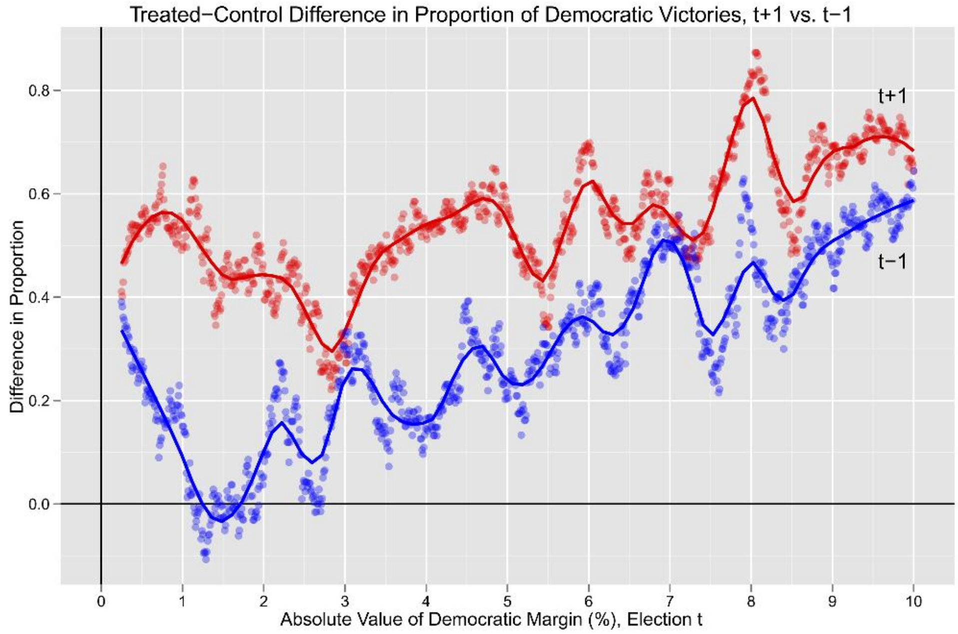


Fig. 3 Divergence of covariate distributions in the immediate neighborhood of the cut-point. Covariate differences are plotted against the midpoint of the disjoint interval tested (e.g., 1.5 for the interval $\{(-1.75, -1.25), (1.75, 1.25)\}$). Loess lines highlight the trends in the imbalance.

in the absolute value of the Democratic margin. The top panel plots the minimum p value of t tests of treated–control differences for 10 covariates.[14] The bottom panel plots the treated–control difference in the proportion of Democratic victories, for $t-1$ and $t+1$. In both panels, covariate balance improves as the interval approaches the threshold until it reaches a margin of 1.5%, at which point covariate imbalance

[14]These covariates (*Previous Democratic Victory*, *Previous Democratic Percent*, *Previous Democratic Margin*, *Democratic Presidential Margin*, *Election Swing*, *Democratic Secretary of State*, *Democratic Governor*, *Election Year*, and dummies for Democratic and Republican incumbent candidates) were selected because data on them are available over the entire time period and throughout the range of the forcing variable.

begins to deteriorate rapidly. The bottom figure shows that even in the interval with the best balance, approximately 1.25%–1.75% away from the cut-point, the treated–control differences in *Democratic Victory* $t+1$ remain large. Balance in the optimal interval is quite good on many important covariates, but among all elections in this interval where data are available, campaign spending is still not equal between winners and losers.

As the divergence near the cut-point indicates, it is not plausible that the imbalances persist merely because nonparametric comparisons ignore trends towards convergence. Rather, the discrepancy between the parametric and nonparametric estimates is primarily due to the fact that Lee's quartic polynomial specification is only moderately sensitive to the behavior of the data in the immediate vicinity of the cut-point. By contrast, alternative specifications such as third- or seven-order polynomials do estimate statistically significant discontinuities. The point is not that the alternative specifications are necessarily correct but rather that we do not have (and Lee does not present) a compelling reason for choosing a quartic polynomial over the other possibilities.[15]

Local linear regression with bandwidths chosen by cross-validation (Imbens and Lemieux 2008) or algorithmic MSE minimization (Imbens and Kalyanaraman 2009) also estimates discontinuities, but these bandwidth selection procedures sometimes chose bandwidths that are at least an order of magnitude larger than the window in which the covariates begin to diverge. As as result, they too may be relatively insensitive to the behavior right around the cut-point and substantially understate the size of the discontinuity. Both methods use parametric approximations and/or data far from the cut-point to inform the choice of bandwidth. The large bandwidths they sometimes select may only be optimal under the assumption that the data-generating process is stable over the range of the data, which we show to be unrealistic in this context.

In summary, bare winners and losers in U.S. House elections differ markedly on many observed covariates, particularly those related to the party of the current incumbent. Incumbent-party candidates generally have more electoral experience and more money than their opponents, and these key variables are also imbalanced between bare winners and losers. Indeed, the outcome of the closest elections is quite accurately predicted by experts' pre-election race ratings. The convergence of covariate distributions presumed by RD reverses abruptly in the immediate neighborhood of the cut-point and becomes particularly severe in elections decided by less than 0.5%. Because the behavior of the data right around the cut-point is so distinctive, parametric and even semi-parametric models that extrapolate to the cut-point will not necessarily detect the sorting at the threshold.

4 Explanations and Substantive Implications

The sudden and dramatic divergence of covariate distributions right around the cut-points is consistent with the well-known fact that highly competitive races are not like normal congressional elections. Voters, the media, candidates, and other strategic actors all behave differently than they do when the outcome is not in doubt. Further, the procedures by which these elections are resolved—such as recounts and legal contests—are also distinctive. In this section, we review potential mechanisms for the sorting in U.S. House elections, offering evidence that postelection manipulation of the vote count (legal or otherwise) is not the primary explanation. Given that candidates in close races are not evenly matched, have quite precise information about the vote as it comes in, and have incentives to make maximal use of their (differential) resources when facing a narrow defeat, sorting is probably the result of activities on or before Election Day.

4.1 *Close Elections Are Not Like Other Elections*

Voters, candidates, political elites, and the media all behave distinctively in close U.S. House elections. A rational calculus of voting suggests that voters' strategic incentives regarding whether and how to vote depends their probability of being pivotal and thus on the closeness of the election (Riker and Ordeshook 1968; Feddersen, Gailmard, and Sandroni 2009). In aggregate terms, voters' utility may in fact

[15] Section 4 of Lee and Lemieux (2010) provides a nuanced and useful discussion of model specification for RD designs, though with an emphasis on estimating treatment effects rather than testing the plausibility of the identifying assumptions.

be maximized by electing incumbents by the narrowest possible margin, thereby retaining good “types” of representatives while preserving the credibility of sanctioning (Fearon 1999). The widely observed empirical regularity of higher turnout in elections expected to be close is the consequence not only of the calculations of voters but also of the strategic activities of political elites (Cox and Munger 1989; Geys 2006, 646–8). Strategic elites direct their scarce resources to the races where they will have the greatest marginal impact. Campaign donations, for example, are concentrated on those expected to be close (Erikson and Palfrey 2000; see also Ansolabehere and Snyder 2000; Glasgow 2002; and Wand 2007). Other targetable resources, such as assistance from campaign consultants and pollsters, are also allocated to competitive races (for an example, see Drucker 2010).

Because “political leaders focus their efforts on the tight contests and forget about the cakewalks,” voters in competitive races are more likely to be mobilized by campaigns (Rosenstone and Hansen 2003, 35). The greater intensity of campaigns in close elections also influences the behavior of social intermediaries like media organizations, who give much more intense coverage to tight U.S. House races than less competitive ones (Jackson 1996; Clarke and Evans 1983, 52). Whether communicated directly by the campaigns or filtered through the media, the heightened flows of information in close elections exert greater persuasion, priming, informing, and mobilizing effects on potential voters (e.g., Basinger and Lavine 2005; Brady, Johnston, and Sides 2006).

Close races are especially likely to elicit distinctive behavior from the candidates themselves, who are both most invested in and best informed about the election. To avert a narrow defeat, candidates may engage in costly activities they would rather avoid, such as making distasteful promises, calling in one-time favors, or even engaging in illegal campaign activities. The incentive for candidates to engage in extraordinary efforts increases with the closeness of the race, and the opportunity to influence the outcome of the contest does not necessary end once the first votes have been cast.

Finally, close elections are distinctive with respect to the procedures by which the ultimate vote total is determined. Because the exact distribution of votes is so consequential, very close elections are frequently subject to retabulations, recounts, and legal challenges. While the most common threshold for triggering a recount is a margin of less than 0.5 percentage points, standards have varied across time and jurisdiction, and whether a recount actually occurs often depends on the discretion of election officials and whether the trailing candidate is willing and able to pay for it (Tokaji and Stoller 2004; Benenson 2010). As is discussed in further detail below, we find that recounts occur in approximately half of all elections decided by less than 0.75%.

4.2 *Explanations for Sorting in Close Elections*

We now turn to a discussion of the plausibility of various explanations for the sorting documented in Section 3. In order for sorting to occur, several conditions must hold. First, one candidate (or her allies) must have very precise information about the vote margin and thus the number of votes needed to sway the outcome. Second, the relevant campaign must have access to superior resources and make maximal use of them only when the outcome is on the line.[16] Thus, the special efforts probably entail costs (in terms of money, political damage, or legal risk) that candidates are not willing to pay except in extremity.

To help adjudicate among competing explanations for sorting, we randomly sampled 75 U.S. House elections in the period 1944–2006 that were decided by a margin of less than 0.75%.[17] With the assistance of a team of research assistants, we collected data on the initial margin reported in each election and determined whether a recount had been conducted. In addition, news reports were used to construct a narrative of each election, with particular attention paid to evidence of vote fraud or other malfeasance.

One potential explanation for sorting is that certain candidates are simply able to steal just enough votes to win. Certainly, vote fraud has a long pedigree in American politics. At numerous points in U.S. history, urban bosses—Chicago’s Daley, Kansas City’s Pendergast, Memphis’s Crump—as well as rural

[16] If candidates made such efforts regardless of the closeness of the election, this would only shift their margin by a constant amount without affecting the dynamics around the threshold.

[17] We sampled in a 0.75% window rather than a 0.5% one in order to determine whether the dynamics of close elections changed at 0.5%, the most common threshold for triggering a recount.

magnates like the Parrs of South Texas and Leander Perez of Louisiana's Plaquemines Parish exercised close control over local election returns, much of it based on electoral manipulation and fraud. One of the cardinal rules of vote fraud is to always wait until all other precincts have reported so you know exactly how many votes you need to manufacture to win (Campbell 2005, 207),[18] so it is easy to imagine how such a practice, if it were widespread, could result in very precise sorting in close elections.

Our evidence, however, suggests that vote fraud is unlikely to be a significant part of the story during our time period.[19] First, in our intensive analysis of a sample of close races, we find that although losing candidates often claim that there were "irregularities" in the election, there are few credible examples of coordinated vote fraud. Of course, since perpetrators of vote fraud are unlikely to advertise their work, definitive evidence is hard to come by. But even scholars convinced that vote rigging was once rampant in American elections tend to believe that it has declined over the last half century with the demise of urban machines and the democratization of rural authoritarian enclaves (Lehoucq 2003, 234).

Modern elections, particularly close ones, are subject to intense scrutiny from media and advocacy groups, and they are increasingly subject to postelection audits and other checks.[20] As Jenkins (2004, 116) demonstrates, the prevalence of contested U.S. House elections, which ranged as high as 10% of all seats in the late-19th century, has averaged below 1% over the past 60 years. Most recent allegations of fraud in House elections, such as in the 1996 Dornan–Sanchez race in California, have been largely refuted (see Minnite 2010, 49–56 and *passim*). Historical works on electoral corruption include almost no examples of fraud having swayed the outcome of a U.S. House election since the 1940s.[21] One possible reason for this is that local areas where machine-style vote fraud is still committed tend not to be in competitive congressional districts and thus only have the potential to affect the outcomes of primary contests or of statewide general elections.

Although new voting technologies have opened new potential opportunities for fraud (Alvarez and Hall 2010, 225–6), the increased scrutiny from media and advocacy groups has not revealed evidence of increased fraud. Therefore, it appears likely that electoral corruption has diminished in the United States, at least since the 1960s. We would therefore expect that if fraud were the culprit, sorting would be less severe in more recent elections. This expectation is not born out in the data: the pro-incumbent party bias in close elections is evident over our entire time period.[22] Thus, although we cannot rule vote fraud out as a potential mechanism, it does not appear to be the most likely explanation for sorting in U.S. House elections, at least in our time period.

A second potential explanation is that some candidates possess systematic advantages in the recount process by which very close elections are often decided. It is reasonable to suspect that candidates with better organization, information, and financing are advantaged in these postelection contests.[23] Indeed, this is the main message of "The Recount Primer" (Downs, Sautter, and Young 1994), a well-known recount handbook for candidates (see Weiner 2010 and Potholm 2003, 94 for similar themes). Although there is some evidence that recounts matter, there is again no smoking gun.

In our random sample of elections in a 0.75% window, we found that at least one vote recount (and sometimes more) occurred in approximately half of the elections. In a number of other elections, the vote was retabulated without an official recount. We find that imbalance on previous Democratic victory is indeed greater in elections in which a recount took place than in non-recounted elections (a treated–control difference of 0.56 versus 0.12). Because candidates are more likely to contest very close

[18] Neglecting this rule is what caused Lyndon Johnson to lose his first close contest for the Senate. As Caro (1990, chap. 13) colorfully relates, Johnson did not repeat the mistake in his second run in 1948, making sure that his supporters in South Texas delayed reporting their (fabricated) results until he knew exactly the margin needed to prevail.

[19] Fraud may, however, have played a major role in sorting near the cut-point in earlier periods in U.S. history (Grimmer et al. 2011).

[20] We thank the editor for his observations on this subject.

[21] Campbell's (2005) history of vote fraud in the United States, which generally plays up the prevalence of fraud, mentions only a single U.S. House race in our time period, a 1988 race in East St. Louis, Illinois, which is not in our data set because it was a special election.

[22] See Fig. 2 in online Appendix D for a plot of pro-incumbent party bias over time.

[23] Lee (2001, 6, footnote 9) actually makes this very point:

> Ironically, the empirical analysis may actually benefit from the fact that these extreme "photo-finish" cases are very rare. It is easy to imagine that if all elections were decided by a handful of votes, many would be contested, and it could be that those candidates who are better at the "post-election" battle—for recounts, for example—may be systematically different, ex ante, from those who lose the "post-election" battle.

elections, however, the closeness of the original count is highly correlated with whether a recount occurred, making it impossible to determine whether the greater imbalance in recounted elections exists independently of the closeness of the margin.

A comparison of the first reported margin in our sample of recounted elections and the ultimate margin in those elections produced no evidence that margin changes tends to benefit certain kinds of candidates. Indeed, of the 35 races in our recount sample in which a recount actually occurred, in only three did the margin actually change sign. This comports with the conventional wisdom that recounts rarely alter election outcomes (Benenson 2010), though even three outcome changes is probably more than one would expect under a model of random and independent misclassification of votes and unbiased recounting (Harris 1988). It is worth noting that in all three races where a recount changed the outcome—as well as in the infamous 1984 McCloskey–McIntyre race in Indiana's 8th district,[24] which was not in our recount sample—the incumbent representative was the ultimate winner. This suggests that when a recount does change the outcome of House election, the incumbent is usually the candidate who benefits.

Neither vote fraud nor recounts appear to account for the sorting in close U.S. House elections. Precise manipulation of the vote margin, however, need not occur after all the votes are cast. In races that go down to the wire, the campaigns intensely monitor the vote as it comes in. The best-organized and most experienced candidates have access to very precise information about how they are performing relative to their opponent. They are even able to react in real time to "random" shocks to their vote share. The effect of a thunderstorm in a friendly town (Lee 2008, 684), for example, can be mitigated by arranging rides to the polls for supporters. Sympathetic judges can be convinced to extend voting hours at particular polling places to compensate for long lines or ballot problems (for an example, see Hauser and Holusha 2006). Last-minute votes can be scrounged up by distributing extra "street money" to ward healers (Beam 2006). Partisan local election officials can exercise discretion over whether to count provisional ballots and other ambiguous votes (Kimball, Kropf, and Battles 2006). At the margins, such actions may systematically push favored candidates over the edge to victory.

We do not claim to have definitely settled the question of how sorting occurs in U.S. House elections, and we encourage future investigation of this subject. Any viable explanation would have to account for the fact that covariate imbalance worsens dramatically right near the cut-point and thus is not the result of the slope in a finite window. The covariates that are most imbalanced are those related to the incumbent party and the candidates' resources and experience, as opposed to the partisanship of state-level officials or the normal presidential vote in the district. Although vote fraud and recounts may contribute, they are not necessary to explain the imbalance. Even as votes are still being cast, candidates have both precise information about the vote margin and the ability to influence it. Further, even in the closest races, candidates are not equally matched in terms of resources and experience, and therefore, some are in a better position to engage in last-minute manipulation.

5 Interpretation and Methodological Implications

In this section, we consider how the existence of sorting around the cut-point affects the interpretation of RD estimates, again using Lee's estimate of the "incumbent party advantage" as a reference point. We evaluate the estimate's sensitivity to hidden bias and also discuss the possibility of covariate adjustment, concluding that potentially strong selection-on-observables and functional-form assumptions are required. We then compare the RD estimate with regression estimates of the "incumbent legislator advantage," showing that while the estimands are conceptually distinct, the RD estimator Lee employs is

[24]This race, the closest in our data set, deviated dramatically from the stylized model in Lee (2008). The contest featured Frank McCloskey, a one-term Democratic incumbent in the U.S. House, and his opponent Rick McIntyre, a Republican state representative. The outcome of this election is actually coded erroneously in Lee's data set as having been won by the challenger McIntyre, when in fact McCloskey was ultimately seated after six months of contestation. On Election Night, the Democrat McCloskey appeared to have survived the Reagan landslide in his district by a mere 74 votes. After the discovery of tabulation errors, however, this margin was revised to a 34-vote deficit in favor of the challenger McIntyre. A partial recount, administered by the Republican secretary of state of Indiana, expanded McIntyre's lead to 418 votes, and the secretary of state quickly certified his victory. The Democratic U.S. House of Representatives refused to seat McIntyre and instead convened a special task force to conduct another recount, which returned a 4-vote victory for McCloskey. When the Democrat was finally seated on May 1, 1985, the House Republican caucus stormed out of the chamber in protest. Charged one Republican representative: "The task force simply found enough votes to elect its man McCloskey and then stopped counting" (Shapiro and Balz, 1985; Herzberg, 1986).

only a simple modification of the commonly used Gelman and King (1990) regression estimator. This comparison demonstrates that the functional-form assumption of the Gelman–King model is stronger in that it pertains to the entire range of the data rather than to the neighborhood of the cut-point only. The Gelman–King estimator also assumes that whether a seat is open in a given election is ignorable. Thus, RD estimates of the incumbency advantage rely on assumptions that are weaker in general than regression methods, though at the cost of changing the estimand.

5.1 *Interpreting the RD Estimate of the "Incumbent Party Advantage"*

The estimand in Lee (2008) is the *incumbent party advantage*: "the overall causal impact of being the current incumbent party in a district on the votes obtained in the district's election" (p. 682). As is discussed in greater detail below, this estimand is conceptually quite different from the *incumbent legislator advantage*, the traditional focus of the incumbency advantage literature in political science. Lee (2008, 686) estimates that Democratic victory in election t increases Democratic vote share in election $t+1$ by over 7 percentage points and Democratic probability of victory by 0.35. A simple comparison of means in a small window around the cut-point yields similar results. These are huge estimated treatment effects, but how should we interpret them in light of the sorting we find at the threshold?

One way to get purchase on this question is to examine the sensitivity of the results to departures from random assignment, according to the method suggested by Rosenbaum (2002). As online Appendix B demonstrates, the odds of treatment assignment would have to differ by a factor of at least $\Gamma = 2.4$ in order for the results to become insignificant. This is a large value of Γ for social science, but as the online Appendix shows formally, it is not implausibly large given the extent of imbalance on even *observed* pretreatment covariates. The variables that are highly imbalanced around the cut-point—incumbent party, financial resources, *CQ* ratings—are plausibly correlated with unobserved confounders such as party organization, candidate skill, and constituency preferences, and it is quite possible that unobserved characteristics are even more imbalanced than the covariates that we can measure. While the true effect may not be 0, it could easily be a fraction of the size of the estimate, an important difference given the degree to which the debate over the incumbency advantage has hinged on whether it has increased over time (e.g., Mayhew 1974; Cox and Katz 1996).

Another potential approach is to adjust for the imbalanced covariates. As Robinson, McNulty, and Krasno (2009, 343) argue, correcting for "some (smallish) level of observed nonequivalence in treatment and control" is a reasonable strategy in many natural experiments. The plausibility of this strategy, however, depends on whether the analyst has controlled for all important determinants of treatment assignment correlated with the outcome (the "selection-on-observables" assumption). In addition, either the covariate distributions must overlap sufficiently or assumptions about functional form are required.

Adjustment may also be motivated if one can plausibly assume that the RD smoothness assumption holds within observed strata—for example, within strata defined by the partisanship of districts. The assignment probabilities across strata may well be different, even at the cut-point, but smoothness will hold within strata. In the case of U.S. elections, however, we find no strata within which the smoothness assumption is plausible. The outcomes of close elections are so predictable that it is impossible to obtain covariate balance between matched treated and control observations. Thus, strong and unverifiable assumptions about the functional form of the relationship between treatment, covariates, and outcome must be made. Although covariate adjustment may be plausible in other applications of RD, it is not in the case of U.S. House elections except under strong assumptions.[25]

As it happens, controlling for covariates using matching or regression sometimes diminishes the size and statistical significance of the RD estimate of the incumbent party advantage but almost never eliminates it entirely.[26] The estimates are also similar in the interval of Democratic margin identified in

[25] An additional complication is that, as Lee and Lemieux (2010, 289) note, in sharp RD designs there is by construction no overlap on the forcing variable, which in the case of elections may itself be an important confounder.

[26] For open seats in the 0.5% window, the estimated effect of *Democratic Victory* t on *Democratic Victory* $t+1$ controlling for *Democratic Margin* t is almost exactly 0.

Section 3 as optimal in terms of covariate balance.[27] These facts offer some reassurance that an incumbent party advantage does indeed exist. Whether covariate adjustment yields an unbiased estimate, however, depends on the plausibility of selection on observables. Imbalanced as they are, observed covariates may be only proxies for more important unobservables like candidate skill or the strength of local party organization. As explained in Section 3.2, incumbents' ability to pull out close victories in bad years for their party could bias the RD estimate upward; alternatively, the disproportionate retirement of incumbents could cause downward bias.

In sum, the estimated effect of party incumbency is usually very large but so is the imbalance between treated and control groups. Even if one were to assume that balancing the *observed* covariates near the cut-point would yield unbiased estimates, the severity of the lack of overlap makes it impossible to adequately condition on observable covariates without functional form assumptions. Further, sensitivity analysis demonstrates that a realistic degree of deviation from random assignment would be sufficient to cast doubt on the results.

Thus, in the case of U.S. House elections, the RD design yields an estimate of the party incumbency advantage whose credibility—since there is sorting near the cut-point—depends on our ability to model this sorting using observed covariates and functional form assumptions. These are the very assumptions which design-based research attempts to avoid. At the same time, this case also illustrates one of the main advantages of RD: because the assumptions of the design have testable implications, we have much better information about whether its estimates are credible in a given context.

5.2 *Comparison of RD and Regression Approaches to the Incumbency Advantage*

Given the problems with applying RD to U.S. House elections, how do RD estimates of the incumbent party advantage compare with traditional regression-based approaches? Since the estimand is different in RD as well as the estimator, we first distinguish conceptually between the incumbent party advantage and various formulations of the traditional incumbent legislator advantage. We then contrast the RD and regression estimators, concluding that if one finds the change in the estimand acceptable, RD estimates are in general superior.

The first work to estimate the size of incumbent House members' advantage over non-incumbents was Erikson (1971), which asks whether "candidates become stronger vote getters once they become incumbents"—that is, "whether a change in the incumbency variable is associated with a subsequent change in the Congressman's electoral margin" (p. 396). One could regard this question as purely descriptive, but Erikson's use of the phrase "effect of incumbency" (p. 395) implies a casual interpretation. Interpreted causally, Erikson's incumbency advantage hypothetically contrasts the performance of a candidate as a (first-term) incumbent with the same candidate's performance as a non-incumbent, with all other characteristics of the candidate held constant.

Gelman and King (1990) shifted the focus of the incumbency advantage literature to a new estimand. These authors define as treatment the incumbent legislator (i.e., winner of the general election in year t) running again in election $t+1$, with open seats as the control condition.[28] This redefinition changed the unit to which treatment is assigned from the candidate to the House seat, changing the relevant counterfactual comparison along with it. The Gelman–King incumbency advantage subsumes several diverse theoretical quantities, including the generic benefits of incumbency (e.g, incumbency as a voting cue; Ferejohn 1977), the "personal vote" (Cain, Ferejohn, and Fiorina 1987), the "scare-off" effect on quality challengers (Cox and Katz 1996), and the higher average quality of incumbents themselves (Zaller 1998). Gelman and King are often considered to be estimating the personal incumbency advantage, but we think calling it the incumbent legislator advantage is more precise.

Gelman and King (1990, 1152) use the following model to estimate the incumbency advantage:

$$\mathbb{E}[V_{t+1}] = \beta_0 + \beta_1 P_t + \beta_2 (P_t \times R_{t+1}) + \beta_3 V_t, \tag{2}$$

[27]Of course, these races are not comparable on a potential crucial confounder: Democratic margin in election t. In addition, the theoretical appeal of the RD design—that treatment assignment changes discontinuously at a sharp observable threshold—is lost in this comparison.

[28]Gelman and Huang (2008) advocate reversing the treatment labels, designating open seats as "treated" rather than "control."

where $V_t \in [0, 1]$ is the *Democratic Share* of the vote in election t, $P_t \in \{-1, 1\}$ is the *Winning Party* in election t, and $R_{t+1} \in \{0, 1\}$ is a dummy variable indicating whether the *Incumbent Runs* in election $t + 1$. Gelman and King characterize β_2 as an unbiased estimator for the incumbency advantage.[29] This claim assumes that, conditional on a linear combination of *Democratic Share t* and *Winning Party t*, the vote share the Democratic candidate would have received in election $t + 1$ is unrelated to whether the incumbent runs in that election—that is, treatment assignment is independent of the potential outcomes.

As noted above, Lee (2008) focuses on the incumbent party advantage as opposed to the traditional concern of the political science literature, the incumbent legislator advantage. In Lee's set-up, the units are House seats; the treatment is victory of the Democratic candidate; and the counterfactual is the difference in Democratic fortunes in election $t+1$ depending on whether the Democratic candidate wins election t in that same district. Conceptually, the two estimands are entirely distinct. Gelman (2011) points out that if House members were allowed to serve only one term, "there would then never be any incumbents running for reelection (thus no incumbency effect) but there would be an incumbent party effect. [The Gelman–King] estimate of incumbency advantage would be undefined, but the Lee estimator of incumbent party effect would work just fine."

Despite the conceptual distance between the Lee and Gelman–King estimands, however, their estimators are in practice very similar. To see the connection between Lee's RD estimator and the Gelman–King regression estimator, first consider the following modification of the Gelman–King model:

$$\mathbb{E}[V_{t+1}] = \delta_0 + \delta_1 P_t + \delta_2(P_t \times R_{t+1}) + \delta_3 M_t + \delta_4 M_t^2 + \delta_5 M_t^3 + \delta_6 M_t^4 + \delta_7(M_t \times P_t) + \delta_8(M_t^2 \times P_t) + \delta_9(M_t^3 \times P_t) + \delta_{10}(M_t^4 \times P_t). \tag{3}$$

Model 3 differs from the Gelman–King estimator (Model 2) in only two respects. First, *Democratic Margin* in election t, denoted M_t, is substituted in place of *Democratic Share* (V_t) on the right-hand side (in two-candidate elections, share and margin are simple transformations of each another). Second, whereas Model 2 has only a linear term, Model 3 includes second-, third-, and fourth-order polynomials in Democratic margin interacted with Democratic victory, making its control for this variable much more flexible. As a substantive matter, these modifications make no difference to estimates of the incumbent legislator advantage (represented by β_2 and δ_2, respectively).

Compare Model 3 with Lee's RD specification, trivially modified so that $P_t \in \{-1, 1\}$ is substituted for Democratic Victory $\in \{0, 1\}$:

$$\mathbb{E}[V_{t+1}] = \gamma_0 + \gamma_1 P_t + \gamma_2 M_t + \gamma_3 M_t^2 + \gamma_4 M_t^3 + \gamma_5 M_t^4 + \gamma_6(M_t \times P_t) + \gamma_7(M_t^2 \times P_t) + \gamma_8(M_t^3 \times P_t) + \gamma_9(M_t^4 \times P_t). \tag{4}$$

The sole difference between Models 3 and 4 is the former's inclusion of an interaction between P_t and R_{t+1}. With the inclusion of this term in Model 3, the coefficient on P_t (δ_1) represents the incumbent party advantage in seats open in election $t + 1$.[30] Gelman and King interpret the coefficient on the interaction term (δ_2) as the incumbent legislator advantage. Alternatively, one could think of it as the difference in the conditional association between party incumbency and Democratic vote share depending on whether the candidate who wins election t runs again in election $t + 1$.

In U.S. House elections 1942–2008,[31] Gelman and King's original specification (2) and our modified version of it (3) result in nearly identical estimates of the incumbent legislator advantage: on average, a party's vote share is about 9% greater in seats where one of its incumbents is running for reelection than in seats where one of its incumbents is retiring. The estimates of β_1 and δ_1 are actually somewhat negative, indicating that conditional on the lagged vote margin, a party performs *better* in open-seat elections where the party *lost* the previous election than in open seats where the party won last time. Lee's incumbent party advantage (γ_1 in Model 4), estimated to be about 9% in this sample, is an average of the party advantage

[29] They treat the incumbency advantage for both parties as symmetric, comparing races with both Democratic and Republican incumbents to open races. Multiplying by P_t merely changes the sign of the term depending on the party of the incumbent.

[30] Since P_t has a two-unit range, its coefficient must be doubled to obtain the difference between Democratic and Republican incumbency.

[31] Elections for which data on the Democratic margin or incumbency status were missing were excluded from this sample.

in non-open seats (+11%) and the party disadvantage in open seats (−7.8%), weighted by the relative frequency of each kind of seat.

Comparing Models 3 and 4 makes clear that if one believes the assumptions of Model 3 (or its simpler counterpart, the original Gelman–King specification in Equation 2), then Lee's estimator is at least as plausible. Both models rely on functional form assumptions. Model 3 (Gelman–King) assumes that incumbent exit $t+1$ is ignorable conditional on a regression model of vote margin t and winning party t. Model 4 (Lee) assumes that the winning party in election t is ignorable (in the limit) conditional on a regression model of vote margin t. The difference is that while the Gelman–King model pertains to the whole range of the data, RD requires only that the model hold near the cut-point if Model 4 is only estimated using such data. Of course, the hope when using RD is that modeling assumptions are not needed—if covariates are balanced near the cut-point, which they are not in the case of U.S. House elections.

Since he examines subsequent districts and not subsequent candidates, Lee's RD estimator has the additional advantage of sidestepping the question of strategic retirement. By contrast, the selection-on-observables assumption of Gelman–King estimator precludes strategic exit of incumbents. The Gelman–King treatment (whether the incumbent runs for reelection) is assigned by the very actor both most invested in the potential outcomes and probably with the best knowledge of them, conditions highly conducive to unobservable selection bias. The RD treatment (which party wins the election), on the other hand, is determined by many actors aside from the incumbent, who may have great influence but probably not complete control over the outcome.[32] Moreover, RD generates empirical implications—covariate balance at the cut-point—that are a consequence of the design rather than the model. It is difficult to imagine a test that would similarly validate Gelman and King's identifying assumption that incumbent exit is exogenous to expected vote share.[33]

Thus, we conclude that although RD does not seem to work in the case of U.S. House elections, its assumptions are still in general weaker than the selection-on-observables and functional-form assumptions required by regression approaches. Although Lee changed the quantity of interest, Erikson and Titiunik (2011) demonstrate in a recent working paper that RD can be used to recover the same estimand as Gelman–King (the incumbent legislator advantage) if one assumes either that there are no strategic retirements (the same assumption as Gelman–King) or if one assumes one can model such behavior. Therefore, one is not forced to change the estimand in order to use RD, although Lee's estimand follows naturally from the design and does not require additional assumptions.

6 Conclusions

A core motivation of the design movement is that without randomization, a discontinuity, or other strong design, no amount of econometric or statistical modeling can make the move from correlation to causation persuasive. When the assignment of treatment is not under the direct control of the researcher, however, it is often difficult to establish that the assignment of treatment is indeed ignorable. In an RD design, it is not *a priori* obvious the degree to which agents have control over the value of the forcing variable near the cut-point. In the case of popular elections, the level of control may vary from one political setting to another. Theory and data must be brought to bear on these questions.

We show that in the case of postwar U.S. House elections, the assumptions of the RD design—most notably, the continuity of potential outcomes—are not satisfied. Bare Democratic victories and bare Democratic defeats differ markedly in terms of the party of the incumbent, national partisan tides, the political

[32] One reasonable question that might arise in response to this characterization is whether the "fuzzy" RD (FRD) design should be applied to U.S. House elections instead of the "sharp" design used by Lee (see Imbens and Lemieux 2008 and Lee and Lemieux 2010 for reviews). In the FRD design, treatment assignment is not fully determined by whether a unit's score on the assignment variable is above the threshold. Rather, treatment assignment is a joint function of this exogenous component and a component that may be endogenous to the outcome of interest. The local average effect of treatment on compliers can be estimated by using the threshold as an instrument. The case of U.S. House elections is superficially similar in that there is substantial but not perfect sorting at the cut-point, indicating that the assignment variable is manipulable but at least some elections may be randomly decided. The obstacle to using the FRD design in this context is that the unmanipulated (random) component of the assignment variable cannot be observed separately from the systematic component and thus cannot be used as an instrument.

[33] See Ansolabehere and Snyder (2004) and Gelman and Huang (2008) for recent defenses of this claim; for critiques, see Cox and Katz (2002) and Katz (2008). Also, recall our finding that about 20% of incumbents who barely win reelection retire in the next election, whereas no victorious challengers in our sample retire.

experience and financial resources of the candidates, and the expectations of well-informed observers. While Lee (2008) is to be commended for checking for discontinuities of the covariates available to him, excessive reliance on parametric extrapolation caused him and other scholars to miss the covariate imbalance, which worsens dramatically in the closest elections.

Given the implausibility of the smoothness assumption in this context, inferences about the causal effect of Democratic victory cannot rest on the design alone; additional assumptions must be made. The sorting in close House elections is so extreme, however, that the functional form and selection-on-observables assumptions required for covariate adjustment to yield unbiased estimates are themselves quite strong. The results are sensitive to a plausible degree of departure from randomness, and the true incumbent party advantage could easily be much smaller than the RD estimate. Nevertheless, a comparison of the Lee RD estimator with traditional regression approaches to the incumbency advantage reveals that RD relies on weaker assumptions, though at the cost of changing the theoretical quantity being estimated.

Our analysis of the explanations for sorting reveals important patterns. We show that whether *ex ante* or *ex post* measures of closeness are used, candidates in highly competitive elections are not equally matched. The incumbent party candidate generally has more money as well as more political experience, contradicting the common intuition that in the closest contests, the candidates will have equal resources and hence an equal chance of winning. Although some electoral corruption and biases in the recount process undoubtedly exist, we present evidence that neither mechanism is the primary explanation for the sorting we document.

Our examination of the dynamics of close elections leaves many questions unanswered, and we hope that our analysis stimulates further research on the subject. One potentially profitable line of inquiry would be a comparison of elections cross-nationally and at different levels of government. As congressional elections have become more professionalized and resource intensive, sorting around the cut-point has, if anything, increased. This trend suggests that close elections in less professionalized settings—state legislative races, for example—may be closer to being decided "as if randomly." Elections for state-wide offices may also exhibit less sorting because on average, they are less influenced by national partisan tides and may require more resources to change the outcomes of close elections. Along the same lines, candidates in local elections in small rural towns may have more control than candidates in large cities in countries where party machines are weak and public opinion polls and other modern methods of measuring public opinion are rare. Finally, the design may perform better in electoral systems where the threshold is more difficult to predict—for example, a plurality voting system with more than two parties.

Our results suggest several concrete recommendations regarding the use of RD designs:

- The burden is on the researcher to provide affirmative evidence for the validity of RD's assumptions in a given context. This requires that she identify and collect accurate data on the observable covariates most likely to reveal sorting at the cut-point. A good rule of thumb is to always check lagged values of the treatment and response variables.
- Careful attention must be paid to the behavior of the data in the *immediate* neighborhood of the cut-point. In this paper, for example, we focus primarily on elections decided by less than 0.5%, a substantially smaller window than many applications consider. This focus reveals that the trend towards convergence evident in wider windows reverses close to the cut-point, a pattern that may occur whenever a highly coveted treatment is assigned via a competitive process with a known threshold.
- Automated bandwidth- and specification-selection algorithms are no sure solution to the problem of modeling data near the cut-point. Even semi-parametric models are based in practice on assumptions about the functional form, sometimes far from the threshold. In our case, for example, the methods recommended in the literature select local linear regression bandwidths that are an order of magnitude larger than the window in which covariate imbalance is most obvious; thus, they are not fully sensitive to this imbalance.

Our analysis shows that it is not a fact of nature that close elections are random. The RD design is a powerful inferential tool that is appropriate in many situations, potentially including many elections. But the applicability of the design cannot be assumed; it must be justified on the basis of context-specific theory and data. Compared to purely observational designs, however, RD's assumptions are clearer and their implications are empirically testable. And although empirical tests reveal that key RD assumptions do not hold in U.S. House elections, the design still makes weaker assumptions than the usual model-based alternatives.

References

Albouy, David. 2009. Partisan representation in Congress and the geographic distribution of federal funds. Working paper no. 15224. National Bureau of Economic Research, Cambridge, MA. http://www.nber.org/papers/w15224.

Alvarez, R. Michael, and Thad E. Hall. 2010. Voting technology. In *The Oxford handbook of American elections and political behavior*, ed. Jan E. Leighley, 219–36. New York: Oxford University Press.

Angrist, Joshua David, and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

———. 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives* 24:3–30.

Ansolabehere, Stephen, and James M. Snyder, Jr. 2000. Soft money, hard money, strong parties. *Columbia Law Review* 100:598–619.

———. 2004. Using term limits to estimate incumbency advantages when officeholders retire strategically. *Legislative Studies Quarterly* 29:488–515.

Basinger, Scott J., and Howard Lavine. 2005. Ambivalence, information, and electoral choice. *American Political Science Review* 99:169–84.

Beam, Christopher. 2006. What's 'Street Money'? Or 'walking-around money'? Or 'get-out-the-vote money'? Slate. 2008. http://www.slate.com/id/2202955/ (accessed 8 May 2011).

Benenson, Bob. 2010. Recount. In *Elections A to Z*. 3rd ed. Washington, DC: CQ Press. http://library.cqpress.com/elections/elaz3d_s472.2.

Benjamini, Y, and Y. Hochberg. 1995. The control of the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* 57:289–300.

Brady, Henry E., Richard Johnston, and John Sides. 2006. The study of political campaigns. In *Capturing campaign effects*, ed. Henry E. Brady and Richard Johnston, 1–26. Ann Arbor: University of Michigan Press.

Brollo, Fernanda, and Tommaso Nannicini. 2010. Tying your enemy's hands in close races: The politics of federal transfers in Brazil. Working paper no. 358, Innocenzo Gasparini Institute for Economic Research, Università Bocconi. http://www.igier.unibocconi.it/folder.php?vedi=4756&tbn=albero&id_folder=182.

Broockman, David E. 2009. Do congressional candidates have reverse coattails? Evidence from a regression discontinuity design. *Political Analysis* 17:418–34.

Butler, Daniel Mark. 2009. A regression discontinuity design analysis of the incumbency advantage and tenure in the U.S. House. *Electoral Studies* 28:123–8.

Butler, Matthew J., and Daniel M. Butler. 2006. Splitting the difference? Causal inference and theories of split-party delegations. *Political Analysis* 14:439–55.

Cain, Bruce, John Ferejohn, and Morris Fiorina. 1987. *The personal vote: Constituency service and electoral independence*. Cambridge, MA: Harvard University Press.

Campbell, Donald T., and Julian C. Stanley. 1963. *Experimental and quasi-experimental designs for research*. New York: Houghton-Mifflin.

Campbell, Tracy. 2005. *Deliver the vote: A history of election fraud, an American political tradition—1742–2004*. New York: Carroll & Graf.

Caro, Robert A. 1990. *Means of ascent*. The Years of Lyndon Johnson. New York: Vintage Books.

Cellini, Stephanie Riegg, Fernando Ferreira, and Jesse Rothstein. 2010. The value of school facility investments: Evidence from a dynamic regression discontinuity design. *Quarterly Journal of Economics* 125:215–61.

Cheng, Ming-Yen, Jianqing Fan, and J. S. Marron. 1997. On automatic boundary corrections. *The Annals of Statistics* 25:1691–708.

Clarke, Peter, and Susan H. Evans. 1983. *Covering campaigns: Journalism in congressional elections*. Stanford, CA: Stanford University Press.

Cook, Thomas D. 2008. 'Waiting for Life to Arrive': A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics* 142:636–54.

Cook, Thomas D., William R. Shadish, and Vivian C. Wong. 2008. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management* 27:724–50.

Cox, Gary W., and Jonathan N. Katz. 1996. Why did the incumbency advantage in U.S. House elections grow? *American Journal of Political Science* 40:478–97.

———. 2002. *Elbridge Gerry's salamander: The electoral consequences of the reapportionment revolution*. New York: Cambridge University Press.

Cox, Gary W., and Mathew D. McCubbins. 2005. *Setting the agenda: Responsible party government in the U.S. House of Representatives*. New York: Cambridge University Press.

Cox, Gary W., and Michael C. Munger. 1989. Closeness, expenditures, and turnout in the 1982 U.S. House Elections. *American Political Science Review* 83:217–31.

DiNardo, John, and David S. Lee. 2004. Economic impacts of new unionization on private sector employers: 1984–2001. *Quarterly Journal of Economics* 119:1383–441.

Downs, Timothy, Chris Sautter, and John Hardin Young. 1994. The Recount Primer. Sautter Communications. http://www.thekingofsteeltown.com/index.html.

Drucker, David M. 2010. Baucus Sent Staff to Aid Reid in Nevada. *Roll Call*, November 5, 2010. http://www.rollcall.com/news/-200258-1.html.

Dunning, Thad. 2008. Improving causal inference: Strengths and limitations of natural experiments. *Political Research Quarterly* 61:282–93.

Eggers, Andrew C., and Jens Hainmueller. 2009. MPs for sale? Returns to office in postwar British politics. *American Political Science Review* 103:513–33.

Erikson, Robert S. 1971. The advantage of incumbency in congressional elections. *Polity* 3:395–405.

Erikson, Robert, and Thomas Palfrey. 2000. Equilibria in campaign spending games: Theory and data. *American Political Science Review* 94:595–609.

Erikson, Robert, and Rocío Titiunik. 2011. Using regression discontinuity to uncover the personal incumbency advantage. Paper presented at the Research Workshop on American Politics, Institute of Governmental Studies, UC Berkeley, May 10, 2011.

Fearon, James D. 1999. Electoral accountability and the control of politicians: Selecting good types versus sanctioning poor performance. In *Democracy, accountability, and representation*, ed. Adam Przeworski, Susan Carol Stokes, and Bernard Manin. New York: Cambridge University Press.

Feddersen, Timothy, Sean Gailmard, and Alvaro Sandroni. 2009. Moral bias in large elections: Theory and experimental evidence. *American Political Science Review* 103:175–92.

Ferejohn, John A. 1977. On the decline of competition in congressional elections. *American Political Science Review* 71:166–76.

Ferreira, Fernando, and Joseph Gyourko. 2009. Do political parties matter? Evidence from U.S. cities. *Quarterly Journal of Economics* 124:399–422.

Fuji, Daisuke, Guido W. Imbens, and Karthik Kalyanaraman. 2009. Notes for Matlab and Stata Regression Discontinuity Software. http://www.economics.harvard.edu/faculty/imbens/files/rd_software_09aug4.pdf. Software (`rdob`, version 10.0).

Gelman, Andrew. 2011. Regression discontinuity designs: looking for the keys under the lamppost? *Statistical Modeling, Causal Inference, and Social Science* (blog), January 15, 2011, http://andrewgelman.com/2011/01/regression_disc/.

Gelman, Andrew, and Gary King. 1990. Estimating incumbency advantage without bias. *American Journal of Political Science* 34:1142–64.

Gelman, Andrew, and Zaiying Huang. 2008. Estimating incumbency advantage and its variation, as an example of a before–after study. *Journal of the American Statistical Association* 103:437–46.

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* 55:326–39.

Geys, Benny. 2006. Explaining voter turnout: A review of aggregate-level research. *Electoral Studies* 25:637–63.

Glasgow, Garrett. 2002. The efficiency of congressional campaign committee contributions in House elections. *Party Politics* 8: 657–72.

Green, Donald P., Terence Y. Leong, Holger L. Kern, Alan S. Gerber, and Christopher W. Larimer. 2009. Testing the accuracy of regression discontinuity analysis using experimental benchmarks. *Political Analysis* 17:400–17.

Grimmer, Justin, Eitan Hersh, Brian Feinstein, and Daniel Carpenter. 2011. 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* 69:201–9.

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* 27:213–27.

Harris, Bernard. 1988. Election recounting. *The American Statistician* 42:66–8.

Hauser, Christine, and John Holusha. 2006. Problems lead 8 states to extend some voting hours. *International Herald Tribune*, November 7, 2006. http://www.nytimes.com/2006/11/08/world/americas/08iht-web.1108day.3441243.html?emc=eta1.

Hays, Jude C., and Robert J. Franzese, Jr. 2007. Estimating the cost of social-democratic government by regression-discontinuity analysis of close elections. Paper presented at the Annual Meeting of the International Political Economy Society, Stanford University, November 2007.

Herzberg, Roberta. 1986. McCloskey versus McIntyre: Implications of contested elections in a federal democracy. *Publius* 16: 93–109.

Horiuchi, Yusaku, and Andrew Leigh. 2009. Estimating incumbency advantage: Evidence from three natural experiments. Paper prepared for presentation at the University of New South Wales, October 2009. http://wwwdocs.fce.unsw.edu.au/economics/news/VisitorSeminar/091014_YusakuHoriuchi.pdf.

ICPSR. 1995. Candidate and constituency statistics of elections in the United States, 1788–1990 computer file, 5th ICPSR edition. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (ICPSR), producer and distributor.

Imbens, Guido W., and Karthik Kalyanaraman. 2009. Optimal bandwidth choice for the regression discontinuity estimator. NBER working paper no. 14726. http://www.economics.harvard.edu/faculty/imbens/files/rd_09feb3.pdf.

Imbens, Guido W., and Thomas Lemieux. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142:615–35.

Jackson, Robert A. 1996. The mobilization of congressional electorates. *Legislative Studies Quarterly* 21:425–45.

Jacob, Suraj, and Naveen Singhal. 2010. Does women's electoral victory increase political participation and success in future elections? Evidence from India. Working paper. http://www.uic.edu/depts/econ/seminar_papers/singhal.pdf.

Jenkins, Jeffery A. 2004. Partisanship and contested election cases in the House of Representatives, 1789–2002. *Studies in American Political Development* 18:112–35.

Katz, Jonathan N. 2008. Comment on 'Estimating Incumbency Advantage and Its Variation, as an Example of a Before–After Study'. *Journal of the American Statistical Association* 103:446–8.

Kimball, David C, Martha Kropf, and Lindsay Battles. 2006. Helping America vote? Election administration, partisanship, and provisional voting in the 2004 election. *Election Law Journal* 5:447–61.

Krasno, Jonathan S. 1994. *Challengers, competition, and reelection: Comparing Senate and House elections*. New Haven, CT: Yale University Press.

Lee, David S. 2001. The electoral advantage to incumbency and voters' valuation of politicians' experience: A regression discontinuity analysis of elections to the U.S. House. Working paper no. W8441, National Bureau of Economic Research. http://emlab.berkeley.edu/users/cle/wp/wp31.pdf.

Lee, David S. 2008. Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics* 142:675–97.

Lee, David S., and Thomas Lemieux. 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* 48:281–355.

Lee, David S., Enrico Moretti, and Matthew J. Butler. 2004. Do voters affect or elect policies? Evidence from the U.S. House. *Quarterly Journal of Economics* 119:807–59.

Lehoucq, Fabrice. 2003. Electoral fraud: Causes, types, and consequences. *Annual Review of Political Science* 6:233–56.

Leigh, Andrew. 2008. Estimating the impact of gubernatorial partisanship on policy settings and economic outcomes: A regression discontinuity approach. *European Journal of Political Economy* 24:256–68.

Listokin, Yair. 2008. Management always wins the close ones. *American Law and Economics Review* 10:159–84.

Mayhew, David R. 1974. Congressional elections: The case of the vanishing marginals. *Polity* 6:295–317.

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

Minnite, Lorraine C. 2010. *The myth of voter fraud*. Ithaca, NY: Cornell University Press.

Pettersson-Lidbom, Per. 2001. Do parties matter for fiscal policy choices? A regression-discontinuity approach. Working paper. http://courses.gov.harvard.edu/gov3009/fall01/Partyeffects.pdf.

Pettersson-Lidbom, Per. 2008. Do parties matter for economic outcomes? A regression-discontinuity approach. *Journal of the European Economic Association* 6:1037–56.

Potholm, Christian P. 2003. *This splendid game: Maine campaigns and elections, 1940–2002*. Lanham, MD: Lexington Books.

Riker, William H. 1962. *The theory of political coalitions*. New Haven, CT: Yale University Press.

Riker, William H., and Peter C. Ordeshook. 1968. A theory of the calculus of voting. *American Political Science Review* 62:25–42.

Robinson, Gregory, John E. McNulty, and Jonathan S. Krasno. 2009. Observing the counterfactual? The search for political experiments in nature. *Political Analysis* 17:341–57.

Rosenbaum, Paul R. 2002. *Observational studies*. 2nd ed. New York: Springer-Verlag.

Rosenstone, Steven J., and John Mark Hansen. 2003. *Mobilization, participation, and democracy in America*. New York: Longman.

Rubin, Donald B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66:688–701.

Schickler, Eric, Kathryn Pearson, and Brian D. Feinstein. 2010. Congressional parties and civil rights politics from 1933 to 1972. *Journal of Politics* 72:672–89.

Shadish, William R., Rodolfo Galindo, Vivian C. Wong, Peter M. Steiner, and Thomas D. Cook. 2011. A randomized experiment comparing random and cutoff-based assignment. *Psychological Methods* 16:179–91.

Shapiro, Margaret, and Dan Balz. 1985. House seats McCloskey. *Washington Post*, May 2, 1985, A1.

Snyder, Jason. 2005. Detecting manipulation in U.S. House elections. Working paper. Haas School of Business, UC Berkeley.

Snyder, Jr., James M. 1990. Campaign contributions as investments: The U.S. House of Representatives, 1980–1986. *Journal of Political Economy* 98:1195–227.

Splawa-Neyman, Jerzy, D. M. Dabrowska, and T. P. Speed. 1923/1990. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science* 5:465–72.

Thistlethwaite, Donald L., and Donald T. Campbell. 1960. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology* 51:309–17.

Titiunik, Rocío. 2009. Incumbency advantage in Brazil: Evidence from municipal Mayor Elections. Working paper. http://www-personal.umich.edu/~titiunik/papers/Titiunik_IABrazil.pdf.

Tokaji, Dan, and Samuel Stoller. 2004. Election Law @ Moritz. Part 5: Voting procedures. Section 5.3—Recounts and ther remedies. http://moritzlaw.osu.edu/electionlaw/ebook/part5/procedures_recount04.html (accessed September 16, 2009).

Trounstine, Jessica. 2011. Evidence of a local incumbency advantage. *Legislative Studies Quarterly* 36:255–80.

Uppal, Yogesh. 2009. The disadvantaged incumbents: Estimating incumbency effects in Indian state legislatures. *Public Choice* 138:9–27.

———. 2010. Estimating incumbency effects in U.S. State Legislatures: A quasi-experimental study. *Economics and Politics* 22:180–99.

Wand, Jonathan. 2007. The allocation of campaign contributions by interest groups and the rise of elite polarization. Working paper. http://wand.stanford.edu/research/mixedmotives.pdf.

Weiner, Jay. 2010. *This is not Florida: How Al Franken won the Minnesota Senate recount*. Minneapolis, MN: University of Minnesota Press.

Zaller, John. 1998. Politicians as prize fighters: Electoral selection and incumbency advantage. In *Party politics and politicians*, ed. John Geer. Baltimore, MD: Johns Hopkins University Press.