

Randomized Experiments from Non-random Selection in U.S. House Elections^{*}

David S. Lee⁺

**Department of Economics
UC Berkeley and NBER**

**(Previous version: September 2003)
January 2005**

Abstract

This paper establishes the relatively weak conditions under which causal inferences from a regression-discontinuity (RD) analysis can be as credible as those from a randomized experiment, and hence under which the validity of the RD design can be tested by examining whether or not there is a discontinuity in any pre-determined (or “baseline”) variables at the RD threshold. Specifically, consider a standard treatment evaluation problem in which treatment is assigned to an individual if and only if $V > v_0$, but where v_0 is a known threshold, and V is observable. V can depend on the individual’s characteristics and choices, but there is also a random chance element: for each individual, there exists a well-defined probability distribution for V . The density function – allowed to differ arbitrarily across the population – is assumed to be continuous. It is formally established that treatment status here is as good as randomized in a local neighborhood of $V = v_0$. These ideas are illustrated in an analysis of U.S. House elections, where the inherent uncertainty in the final vote count is plausible, which would imply that the party that wins is essentially randomized among elections decided by a narrow margin. The evidence is consistent with this prediction, which is then used to generate “near-experimental” causal estimates of the electoral advantage to incumbency.

* An earlier draft of this paper, “The Electoral Advantage to Incumbency and Voters’ Valuation of Politicians’ Experience: A Regression Discontinuity Analysis of Elections to the U.S. House”, is available online as NBER working paper #8441. Matthew Butler provided outstanding research assistance. I thank John DiNardo and David Card for numerous invaluable discussions, and Josh Angrist, Jeff Kling, Jack Porter, Larry Katz, Ted Miguel, and Ed Glaeser for detailed comments on an earlier draft. I also thank seminar participants at Harvard, Brown, UIUC, UW-Madison and Berkeley, and Jim Robinson for their additional useful suggestions.

⁺ Department of Economics, 549 Evans Hall, #3880, Berkeley, CA 94720-3880. dslee@econ.berkeley.edu

1 Introduction

There is a recent renewed interest in the identification issues involved in (Hahn, Todd, and van der Klaauw, 2001), the estimation of (Porter, 2003), and the application of (Angrist and Lavy, 1999; van der Klaauw 2002) Thistlethwaite and Campbell's (1960) regression-discontinuity design (RDD). RD designs involve a dichotomous treatment that is a deterministic function of an single, observed, continuous covariate (henceforth, "score"). Treatment is assigned to those individuals whose score crosses a known threshold. Hahn, Todd, and van der Klaauw (2001) formally establish minimal continuity assumptions for identifying treatment effects in the RDD: essentially, the average outcome for individuals marginally below the threshold must represent a valid counterfactual for the treated group just above the threshold. For the applied researcher, there are two limitations to invoking this assumption: 1) in many contexts, individuals have some influence over their score, in which case it is unclear whether or not such an assumption is plausible, and 2) it is a fundamentally untestable assumption.

This paper describes a very general treatment assignment selection model that 1) allows individuals to influence their own score in a very unrestrictive way, and 2) generates strong testable predictions that can be used to assess the validity of the RDD. In particular, it is shown below that causal inferences from RD designs can sometimes be as credible as those drawn from a randomized experiment.

Consider the following general mechanism for treatment assignment. Each individual is assigned a score V , which is influenced partially by 1) the individual's attributes and actions, and 2) by random chance. Suppose that conditional on the individual's choices and characteristics, the probability density of V is continuous. Treatment is given to the individual if and only if V is greater than a known threshold v_0 . Note that there is unrestricted heterogeneity in the density function for V across individuals, so that each individual will in general have a different (and unobserved to the analyst) probability of treatment assignment.

Below it is formally established that this mechanism not only satisfies the minimal assumptions for RD designs outlined in Hahn, Todd, and van der Klaauw (2001); it additionally generates variation in

treatment status that is as good as randomized by an experiment – in a neighborhood of $V = v_0$. Close to this threshold, *all variables* determined prior to assignment will be independent of treatment status. Thus – as in a randomized experiment – differences in post-assignment outcomes will not be confounded by omitted variables, whether observable or unobservable.

This alternative formulation of a valid RD design and the local independence result are useful for three different reasons. First, it illustrates that natural randomized experiments can be isolated even when treatment status is driven by non-random self-selection. For example, the vote share V obtained by a political candidate could be dependent on her political experience and campaigning effort, so that on average, those who receive the treatment of winning the election ($V > \frac{1}{2}$) are systematically more experienced and more ambitious. Even in this situation, provided that there is a random chance error component to V that has continuous pdf, treatment status in a neighborhood of $V = \frac{1}{2}$ is statistically randomized.

Second, in any given applied context, it is arguably easy to judge whether or not the key condition (continuous density of V for each individual) holds. This is because the condition is directly related to individuals’ incentives and ability to sort around the threshold v_0 . As discussed below, if individuals have exact control over their own value of V , the density for each individual is likely to be discontinuous. When this is the case, the RDD is likely to yield biased impact estimates.

Finally, and perhaps most importantly, the local independence result implies a strong empirical test of the internal validity of the RDD. In a neighborhood of v_0 , treated and control groups should possess the same distribution of baseline characteristics. The applied researcher can therefore verify – as in a randomized controlled trial – whether or not the randomization “worked”, by examining whether there are treatment-control differences in baseline covariates.¹ These specification tests are not based on additional assumptions; rather, they are auxiliary predictions – consequences of the assignment mechanism described above. The local random assignment result also gives a theoretical justification for expecting impact esti-

¹ Such specification checks have been used recently, for example, in Lee, Moretti, and Butler (2004), Linden (2004), Martorell (2004), Clark (2004), Matsudaira (2004), DiNardo and Lee (2004).

mates to be insensitive to the inclusion of any combination of baseline covariates in the analysis.²

The result is applied to an analysis of the incumbency advantage in elections to the United States House of Representatives. It is plausible that the exact vote count in large elections, while influenced by political actors in a non-random way, is also partially determined by chance beyond any actor's control. Even on the day of an election, there is inherent uncertainty about the precise and final vote count. In light of this uncertainty, the local independence result predicts that the districts where a party's candidate just *barely* won an election – and hence barely became the incumbent – are likely to be comparable in all other ways to districts where the party's candidate just *barely* lost the election. Differences in the electoral success between these two groups in the next election thus identifies the causal party incumbency advantage.

Results from data on elections to the United States House of Representatives (1946-1998) yields the following findings. First, the evidence is consistent with the strong predictions of local random assignment of incumbency status around the 50 percent vote share threshold. Among close electoral races, the districts where a party wins or loses are similar along *ex ante*, pre-determined characteristics. Second, party incumbency is found to have a significant *causal* effect on the probability that a political party will retain the district's seat in the next Congress; it increases the probability on the order of 0.40 to 0.45.³ The magnitude of the effect on the vote share is about 0.08. Second, losing an election reduces the probability of a candidate running again for office by about 0.43, consistent with an enormous deterrence effect.

Section 2 provides a brief background on regression-discontinuity designs, reviews the key statistical properties and implications of truly randomized experiments, and formally establishes how the treatment assignment mechanism described above can share those properties. Section 3 describes the inference problem, data issues, and the empirical results of an RDD analysis of the incumbency advantage in the U.S. House. Section 4 concludes.

² Hahn, Todd, and van der Klauww (2001) do state that the “advantage of the method is that it bypasses many of the questions concerning model specification: both the question of which variables to include in the model for outcomes,” but provide no justification for why the treatment effect estimates should be insensitive to the inclusion of baseline characteristics.

³ As discussed below, the causal effect for the *individual* that I consider is the effect on the probability of both becoming a candidate *and* winning the subsequent election. Below I discuss the inherent difficulty in isolating the causal effect *conditional on running for re-election*.

2 Random assignment from non-random selection

In a regression-discontinuity design (RDD) the researcher knows that treatment is given to individuals if and only if an observed covariate V crosses a known threshold v_0 .⁴ In Thistlethwaite and Campbell's (1960) original application of the RDD, an award was given to students who obtained a minimum score on a scholarship examination. OLS was used to estimate differences in future academic outcomes between the students who scored just above and below the passing threshold. This discontinuity gap was attributed to the effect of the test-based award.

Hahn, Todd, and van der Klaauw (2001) was the first to link the RDD to the treatment effects literature, and to formally explore the sources of identification that underlie the research design. There, it is established that the mere treatment assignment rule itself is insufficient to identify any average treatment effect. Identification relies on the assumption that

$$E[Y_0|V = v] \text{ and } E[Y_1|V = v] \text{ are continuous in } v \text{ at } v_0 \quad (1)$$

where Y_1 and Y_0 denote the potential outcomes under the treatment and control states, and V is the score that determines treatment.⁵ This makes clear that the credibility of RDD impact estimates depends on whether or not the mean outcome for individuals marginally below the threshold identifies the true counterfactual for those marginally above the threshold v_0 .

For empirical researchers, however, there are two practical limitations to the assumption in (1). First, in many real-world contexts, it is difficult to determine whether the assumption is plausible. This is because (1) is not a description of a treatment-assigning process; instead, it is a statement of what must be mathematically true if the RD gap indeed identifies a causal parameter. For example, in Thistlethwaite and Campbell's (1960) example, if the RD gap represents a causal effect, then the outcomes for students who barely fail must represent what would have happened to the marginal winners had they not received the scholarship. But at first glance, there appears to be nothing about this context would lead us to believe – or

⁴ More generally, there are two types of designs: the so-called “sharp” and “fuzzy” designs, as described in Hahn, Todd, and van der Klaauw (2001). This paper focuses on the sharp RD.

⁵ This is a simplified re-statement of Assumptions (A1) and (A2) in Hahn, Todd, and van der Klaauw (2001).

disbelieve – that (1) actually holds. Second – perhaps more importantly – assumption (1) is fundamentally untestable; there is no way for a researcher to empirically assess its plausibility.

The discussion below attempts to address these two limitations. It is shown that a somewhat unrestricted treatment-assignment mechanism not only satisfies 1, but the variation in the treatment – in a neighborhood of v_0 – shares the same statistical properties as a classical randomized experiment. As discussed in Section 2.4, the key condition for this result is intuitive and its plausibility is arguably easier to assess than (1) in an applied setting. The plausibility of the key condition is directly linked to how much control individuals have over the determination of the score V . Indeed, it becomes clear how economic behavior can sometimes *invalidate* RDD inferences. Furthermore, as shown in Section (2.2) the “local randomization” result implies that these key conditions generate strong testable restrictions that are analogous to those implied by a true randomized experiment.

2.1 Review of Classical Randomized Experiments

In order to introduce notation and provide a simple basis for comparison, this section formally reviews the statistical properties and implications of classical randomized experiments. The next section will describe a general non-experimental (and non-randomized) treatment assignment mechanism that nevertheless shares these properties and implications – among individuals with realized scores close to the RD threshold.

Consider the following stochastic mechanism: 1) randomly draw an individual from a population of individuals, 2) assign treatment to the individual with constant probability p_0 , and 3) measure all variables, including the outcome of interest. Formally, let (Y, X, D) be observable random variables generated by this process, where Y is the outcome variable of interest, X is any “pre-determined” variable (one whose value has already been determined prior to treatment assignment), and D an indicator variable for treatment status.

Adopting the potential outcomes framework, we imagine that the assignment mechanism above actually generates (Y_1, Y_0, X, D) where Y_1 and Y_0 are the outcomes that will occur if the individual receives or is denied treatment, respectively. For any one individual, we cannot observe Y_1 and Y_0 simultaneously.

Instead, we observe $Y = DY_1 + (1 - D)Y_0$.

To emphasize the distinction between the random process that draws an individual from the population and that which assigns treatment – and to help describe the results in a later section – it is helpful to provide an equivalent description of the data generating process.

Condition 1a. Let (W, D) be a pair of random variables (with W unobservable), and let $Y_1 \equiv y_1(W)$, $Y_0 \equiv y_0(W)$, $X \equiv x(W)$, where $y_1(\cdot)$, $y_0(\cdot)$, and $x(\cdot)$ are real-valued functions.⁶

One can think of W as either the “type” or “identity” of the randomly drawn individual. There is no loss of generality in assuming that it is a one-dimensional random variable; the appendix provides statements of all propositions and their proofs within a measure-theoretic framework. By definition, D is not an argument of either y_1 or y_0 , and since X has already been determined prior to treatment assignment, D is also not an argument of the function x .

Under random assignment, every individual has the same probability of receiving the treatment, so that we have

Condition 2a. $\Pr[D = 1|W = w] = p_0$ for all w in the support of W

As a result, we obtain three well-known and useful implications of a randomized experiment, summarized as follows:

Proposition 1 *If Conditions 1a and 2a hold, then:*

a)

$$\Pr[W \leq w|D = 1] = \Pr[W \leq w|D = 0] = \Pr[W \leq w], \forall w \text{ in the support of } W$$

b)

$$E[Y|D = 1] - E[Y|D = 0] = E[Y_1 - Y_0] = ATE$$

c)

$$\Pr[X \leq x_0|D = 1] = \Pr[X \leq x_0|D = 0], \forall x_0$$

It is easy to see that a) simply follows from Condition 1a and Bayes’ rule. Since the distribution

⁶ The functions must be measurable \mathcal{R}^1 , the class of linear Borel sets.

of W is identical irrespective of treatment status and Y_1 , Y_0 , and X are functions of W , b) and c) naturally follow.

b) is simply a formal statement of the known fact that in a classical randomized experiment, the difference in the conditional means of Y will identify the average treatment effect (ATE). c) is a formal statement of another important consequence of random assignment. It states that *any* variable that is determined prior to the random assignment will have the same distribution in either the treatment or control state. This formalizes why analysts expect predetermined (or “baseline”) characteristics to be similar in the treatment and control groups (apart from sampling variability).

Indeed, in practice, analyses of randomized experiments typically begin with an assessment of the comparability of treated and control groups in the baseline characteristics X . Thus, Condition 2a generates many testable restrictions, and applied researchers find those tests useful for empirically assessing the validity of the assumption.

2.2 Random Assignment from a Regression Discontinuity Design

In most applied contexts, researchers know that assignment to treatment is not randomized as in an experiment. Instead, they believe in non-random self-selection into treatment status. It is shown here that even when this is the case, the RDD can nevertheless sometimes identify impact estimates that share the same validity as those available from a randomized experiment.

Consider the following data generating process: 1) randomly draw an individual from a population of individuals, after they have made their optimizing decisions, 2) assign a score V , drawn from a non-degenerate, sufficiently “smooth” individual-specific probability distribution, 3) assign treatment status based on the rule $D = 1[V \geq 0]$ where $1[\cdot]$ is an indicator function, and 4) measure all variables, including the outcome of interest.

More formally, we have

Condition 1b. Let (W, V) be a pair of random variables (with W unobservable, V observable),

and let $Y_1 \equiv y_1(W)$, $Y_0 \equiv y_0(W)$, $X \equiv x(W)$, where $y_1(\cdot)$, $y_0(\cdot)$, and $x(\cdot)$ are real-valued functions. Also, let $D = 1[V \geq 0]$. Let $G(\cdot)$ be the marginal cdf of W .

Condition 2b. $F(v|w)$, the cdf of V conditional on W , is such that $0 < F(v|w) < 1$, and is continuously differentiable in v at $v = 0$, for each w in the support of W . Let $f(\cdot)$ and $f(\cdot|\cdot)$ be the marginal density of V and the density of V conditional on W , respectively.

Note that by allowing the distribution of V conditional on W to depend on w in a very general way, individuals can take action to influence their probability of treatment. But V has some random chance element to it, so that each individual's probability of receiving treatment is somewhere between 0 and 1. In addition, Condition 2b implies that for each individual, the probability of obtaining a V just below and just above 0 are the same. Note that Condition 2b still allows arbitrary correlation – in the overall population – between V and any one of Y_1 , Y_0 , or X .

The main result is a proposition analogous to Proposition 1:

Proposition 2 *If Conditions 1b and 2b hold, then:*

a)

$$\Pr[W \leq w|V = v], \text{ is continuous in } v \text{ at } v = 0, \forall w$$

b)

$$\begin{aligned} E[Y|V = 0] - \lim_{\Delta \rightarrow 0^-} E[Y|V = \Delta] &= E[Y_1 - Y_0|V = 0] \\ &= \int_{-\infty}^{\infty} (y_1(w) - y_0(w)) \frac{f(0|w)}{f(0)} dG(w) \\ &= ATE^* \end{aligned}$$

c)

$$\Pr[X \leq x_0|V = v], \text{ is continuous in } v \text{ at } v = 0, \forall x_0$$

a), b), and c) are analogous to a), b), and c) in Proposition 1. a) states that the probability distribution of the identity or “type” of individuals is the same just above and below $v = 0$. b) states that the discontinuity in the conditional expectation function identifies an average treatment effect, and c) states that all pre-determined characteristics should have the same distribution just below and above the threshold. c)

implies that empirical researchers can empirically assess the validity of their RDD, by examining whether or not, for example, the mean of *any* pre-determined X conditional on V changes discontinuously around 0. If it does, either Condition 1b or 2b must not hold.

It is important to note that ATE^* is a particular kind of average treatment effect. It is clearly not the average treatment effect for the entire population. Instead, b) states that it can be interpreted as a *weighted* average treatment effect: those individuals who are more likely to obtain a draw of V near 0 receive more weight than those who are unlikely to obtain such a draw. Thus, with this treatment-assignment mechanism, it is misleading to state that the discontinuity gap identifies an average treatment effect “only for the subpopulation for whom $V = 0$ ”, which is, after all, a measure zero event. It is more accurate to say that it is a weighted average treatment effect for the *entire* population, where the weights are the probability that the individual draws a V “near” 0.

2.3 Allowing for the Impact of V

There are two shortcomings to the treatment-assignment mechanism described by Conditions 1b and 2b. First, it may be too restrictive for some applied contexts. In particular, it assumes that the random draw of V does not itself have an impact on the outcome – except through its impact on treatment status. That is, while V is allowed to be correlated with Y_1 or Y_0 in the population, V is not permitted to have an independent causal impact on Y for a given individual. In a non-experimental setting, this may be unjustifiable. For example, a student’s score on a scholarship examination might itself have an impact on later-life outcomes, quite independently of the receipt of the scholarship.

Second, the counterfactuals Y_1 and Y_0 may not even be well-defined for certain values of V . For example, suppose a merit-based scholarship is awarded to a student solely on the basis of scoring 70 percent or higher on a particular examination. What would it mean to receive a test-based scholarship even while scoring 50 on the test, or to be denied the scholarship even after scoring a 90? In such cases, Y_1 is simply not defined for those with $V < 0$, and Y_0 is not defined for those with $V_i \geq 0$. It may nevertheless be of interest to know the direct impact of winning a test-based scholarship on future academic outcomes.

As another example, suppose we are interested in the causal impact of a Democratic electoral victory in a U.S. Congressional District race on the probability of future Democratic electoral success. We know that a Democratic electoral victory is a deterministic function of the vote share. Again, the counterfactual notation is awkward, since it makes little sense to conceive of the potential outcome of a Democrat who lost the election with 90 percent of the vote.

To address the limitations above consider the alternative assumption:

Condition 1c. Let (W, V) be a pair of random variables (with W unobservable, V observable), and let $Y \equiv y(W, V)$, and $X \equiv x(W)$, where for each w , $y(\cdot, \cdot)$ is continuous in the second argument except at $V = 0$, where the function is only continuous from the right. Define the function $y^-(w) = \lim_{\varepsilon \rightarrow 0^+} y(w, \varepsilon)$ and $y^+(w) = y(w, 0)$.

$y(\cdot, \cdot)$ is a response function relating the outcome to a realization of V . For individual w with realization v of the score V , the outcome would be $y(w, v)$. The function $y(\cdot, \cdot)$ is simply an analogue to the potential outcomes “function” utilized in Conditions 1a and 1b, except that the second argument is a continuous rather than a discrete variable. For each individual w , there exists an impact of interest, $y^+(w) - y^-(w)$, and the RD analysis identifies an average of these impacts.

This leads to:

Proposition 3 *If Conditions 1c and 2b hold, then a) and c) of Proposition 2 holds, and:*
b)

$$\begin{aligned} & E[Y|V = 0] - \lim_{\Delta \rightarrow 0^-} E[Y|V = \Delta] \\ &= \int_{-\infty}^{\infty} (y^+(w) - y^-(w)) \frac{f(0|w)}{f(0)} dG(w) \\ &= ATE^{**} \end{aligned}$$

So ATE^{**} is a weighted average of individual-specific discontinuity gaps $y^+(\cdot) - y^-(\cdot)$ where the weights are the same as in Proposition 2.

2.4 Self-selection and Random Chance

The continuity Condition 2b is crucial to the local random assignment results of Proposition 2 and 3. It is

easy to see that if, for a nontrivial fraction of the population, the density of V is discontinuous at the cutoff point, then a), b), and c) of Propositions 2 and 3 will generally not be true. Condition 2b is also somewhat intuitive and its plausibility is arguably easier to assess than 1. Indeed, there is a link between Condition 2b and the ability of agents to manipulate V , particularly around the discontinuity threshold. When agents can precisely manipulate their own value of V , it is possible that Condition 2b will not hold, and the RDD could then lead to biased impact estimates.

For example, suppose a nontrivial fraction of students taking the examination knew with certainty, for each question, whether or not their answer was correct – even while taking the exam. If these students cared only about winning the scholarship per se, and if spending time taking the exam is costly, they would choose to answer the minimum number of questions correctly (e.g. 70) to obtain the scholarship. In this scenario, clearly the density of V would be discontinuous at the cutoff point, and thus the use of the RDD would be inappropriate.

Alternatively, suppose for each student, there is an element of chance that determines the score. The student may not know the answers to all potential questions, so that at the outset of the examination, which of those questions will appear has a random component to it. The student may feel exceptionally sharp that day or instead may have a bad “test” day, both of which are beyond the control of the student. If this is a more believable description of the treatment assignment process, then Condition 2b would seem plausible.

One way to formalize the difference between these two different scenarios is to consider that V is the sum of two components: $V = Z + e$. Z denotes the systematic, or predictable component of V that can depend on the individuals’ attributes and/or actions (e.g. students’ efforts in studying for the exam), and e is an exogenous, random chance component (e.g. whether the “right” questions appear on the exam, having a good “testing” day), with a continuous density. In the first scenario, there was no stochastic component e , since the student knew exactly whether each of his answers was correct. In the second scenario, however minimally, the component e – random chance – does influence the final score V .

In summary, Propositions 2 and 3 show that localized random assignment can occur even in the

presence of endogenous sorting, as long as agents do not have the ability to sort precisely around the threshold. If they can, the density of V is likely to be discontinuous, especially if there are benefits to receiving the treatment. If they cannot – perhaps because there is ultimately some unpredictable, and uncontrollable (from the point of view of the individual) component to V – the continuity of the density may be justifiable.

2.5 Relation to Selection Models

The treatment-assignment mechanism described by Conditions 1b (or 1c) and 2b has some generality. The conditions are implicitly met in typical econometric models for evaluation studies (except for the observability of V). For example, consider the reduced-form formulation of Heckman’s (1978) dummy endogenous-variable model:

$$\begin{aligned} y_1 &= x_1\alpha_1 + d\beta + u_1 \\ y_2^* &= x_2\alpha_2 + u_2 \\ d &= 1 \text{ if } y_2^* \geq 0 \\ &= 0 \text{ if } y_2^* < 0 \end{aligned} \tag{2}$$

where y_1 is the outcome of interest, d is the treatment indicator, x_1 and x_2 are exogenous variables and (u_1, u_2) are error terms that are typically assumed to be bivariate normal and jointly independent of x_1 and x_2 . An exclusion restriction typically dictates that x_2 contains some variables that do not appear in x_1 .

Letting $V = y_2^*$, and $Y_1 = x_1\alpha_1 + \beta + u$, $Y_0 = x_1\alpha_1 + u$, and $D = d$, it is clear that this conventional selection model satisfies Conditions 1b and 2b, except that y_2^* here is unobservable. In this setting, it is crucial that the specification (e.g. the choice of variables x_1 and x_2 , the independence assumption, the exclusion restriction) of the model is correct. Any mis-specification (e.g. missing some variables, correlation between the errors and x_1 and x_2 , violation of exclusion restriction) will lead to biased estimates of α_1 , α_2 , and β .

When, on the other hand, the researcher is fortunate enough to directly observe y_2^* – as in the RDD

– none of the variables in x_1 or x_2 are needed for the estimation of β . And it is also unnecessary to assume independence of the errors u_1, u_2 . If x_1 and x_2 are available to the researcher (and insofar as they are known to have been determined prior to the assignment of d), they can be used to check the validity of the continuity Condition 2b, which drives the local random assignment result. Propositions 2 and 3 imply that this can be done, for example, by examining the difference $E[x_1|y_2^* = 0] - \lim_{\Delta \rightarrow 0^-} E[x_1|y_2^* = \Delta]$. If the local random assignment result holds, this difference should be zero. The variables x_1 and x_2 serve another purpose in this situation. They can be included in a regression analysis to reduce sampling variability in the impact estimates. Local independence implies that the inclusion of those covariates will lead to alternative, consistent estimates, with generally smaller sampling variability. This is analogous to including baseline characteristics in the analysis of randomized experiments.

It should be noted that this connection between RDD and selection models is not specific to the well-known parametric version of Equation 2. The arguments can easily be extended for a more generalized selection model that does not assume, for example, the linearity of the indices $x_1\alpha_1$ or $x_2\alpha_2$, the joint normality of the errors, or the implied constant treatment effect assumption. Indeed, Condition 1b (or 1c) is perhaps the least restrictive description possible for a selection model for the treatment evaluation problem.

3 RDD analysis of the Incumbency Advantage in the U.S. House

This section applies the ideas developed above to the problem of measuring the electoral advantage of incumbency in the United States House of Representatives. In the discussion that follows, the “incumbency advantage” is defined as the overall causal impact of being the current incumbent party in a district on the votes obtained in the district’s election. Therefore, the unit of observation is the Congressional district. The relation between this definition and others commonly used in the political science literature is discussed briefly in Section 3.5 and in more detail in Appendix B.

3.1 The Inference Problem in Measuring the Incumbency Advantage

One of the most striking facts of congressional politics in the United States is the consistently high rate of electoral success of incumbents, and the electoral advantage of incumbency is one of the most studied

aspects of research on elections to the U.S. House [Gelman and King, 1990]. For the U.S. House of Representatives, in any given election year, the incumbent party in a given congressional district will likely win. The solid line in Figure I shows that this re-election rate is about 90 percent and has been fairly stable over the past 50 years.⁷ Well-known in the political science literature, the electoral success of the incumbent party is also reflected in the two-party vote share, which is about 60 to 70 percent during the same period.⁸

As might be expected, incumbent *candidates* also enjoy a high electoral success rate. Figure I shows that the winning candidate has typically had an 80 percent chance of both running for re-election and ultimately winning. This is slightly lower, because the probability that an incumbent will be a candidate in the next election is about 88 percent, and the probability of winning, conditional on running for election is about 90 percent. By contrast, the runner-up candidate typically had a 3 percent chance of becoming a candidate and winning the next election. The probability that the runner-up even becomes a candidate in the next election is about 20 percent during this period.

The overwhelming success of House incumbents draws public attention whenever concerns arise that Representatives are using the privileges and resources of office to gain an “unfair” advantage over potential challengers. Indeed, the casual observer is tempted to interpret Figure I as evidence that there is an electoral *advantage* to incumbency – that winning has a *causal* influence on the probability that the candidate will run for office again and eventually win the next election. It is well-known, however, that the simple comparison of incumbent and non-incumbent electoral outcomes does not necessarily represent anything about a true electoral advantage of being an incumbent.

As is well-articulated in Erikson [1971], the inference problem involves the possibility of a “reciprocal causal relationship”. Some – potentially *all* – of the difference is due to a simple selection effect: incumbents are, by definition, those politicians who were successful in the previous election. If what makes them successful is somewhat persistent over time, they should be *expected* to be somewhat more successful when running for re-election.

⁷ Calculated from data on historical election returns from ICPSR study 7757. See Data Appendix for details. Note that the “incumbent party” is undefined for years that end with ‘2’ due to decennial congressional re-districting.

⁸ See, for example, the overview in Jacobson [1997].

3.2 Model

The ideal thought experiment for measuring the incumbency advantage would exogenously change the incumbent party in a district from, for example, Republican to Democrat, while keeping all other factors constant. The corresponding increase in Democratic electoral success in the next election would represent the overall electoral benefit due to being the incumbent party in the district.

There is an RDD inherent in the U.S. Congressional electoral system. Whether or not the Democrats are the incumbent party in a Congressional district is a deterministic function of their vote share in the prior election.

Assuming that there are two parties, consider the following model of Congressional elections:

$$v_{i2} = \alpha w_{i1} + \beta v_{i1} + \gamma d_{i2} + e_{i2} \quad (3)$$

$$d_{i2} = 1 \left[v_{i1} \geq \frac{1}{2} \right]$$

$f_{i1}(v|w)$ – density of v_{i1} conditional on w_{i1} – is continuous in v

$$E[e_{i2}|w_{i1}, v_{i1}] = 0$$

where v_{it} is the vote share for the Democratic candidate in Congressional district i in election year t . d_{i2} is the indicator variable for whether the Democrats are the incumbent party during the electoral race in year 2. It is a deterministic function of whether the Democrats won election 1. w_{i1} is a vector of variables that reflect all characteristics determined or agents' choices as of election day in year 1.

The first line in (3) is a standard regression model describing the causal impacts of w_{i1} , v_{i1} , and d_{i2} on v_{i2} . w_{i1} could represent the partisan make-up of the district, party resources, or the quality of potential nominees. v_{i1} is also permitted to impact v_{i2} . For example, a higher vote share may attract more campaign donors, which in turn, could boost the vote share in election year 2. The potentially discontinuous jump in how v_{i1} impacts v_{i2} is captured by the coefficient γ , and is the parameter of interest – the electoral advantage to incumbency.

The main problem is that elements of w_{i1} may be unobservable to the researcher, so OLS will

suffer from an omitted variables bias, since w_{i1} might be correlated with v_{i1} , and hence with d_{i2} . That is, the inherent advantages the Democrats have in a congressional district (e.g. the degree of liberalness of the constituency, party resources allocated to the district) will naturally be correlated with their electoral success in year 1, and hence will be correlated with whether they are the incumbent party during the electoral race in year 2. This is why a simple comparison of electoral success in year 2, between those districts where the Democrats won and lost in year 1, is likely to be biased.

But an RDD can plausibly be used here. Letting $W = w_{i1}$, $V = v_{i1}$, and $Y = y(W, V) = \alpha W + \beta V + \gamma \cdot 1[V \geq \frac{1}{2}]$, we have $\gamma = y(w, \frac{1}{2}) - \lim_{\varepsilon \rightarrow 0^+} y(w, \frac{1}{2} - \varepsilon)$. Conditions 1c and 2b hold, and so Proposition 3 applies.⁹ Intuitively, conditional on agents' actions and characteristics as of election day, if there exists a random chance element (that has a continuous density) to the final vote share v_{i1} , then whether the Democrats win in a closely-contested election is would determined as if by a flip of a coin. As a consequence, we can obtain credible estimates of the electoral advantage to incumbency by comparing the average Democratic vote shares in year 2 between districts in which Democrats narrowly won and narrowly lost elections in year 1.

The crucial assumption here is that – even if agents can influence the vote – there is nonetheless a non-trivial random chance component to the ultimate vote share, and that conditional on the agents' choices and characteristics, the vote share v_{i1} has a continuous density. It is plausible that there is at least some random chance element to the precise vote share. For example, the weather on election day can influence turnout among voters.

Assuming a continuous density requires that certain kinds of electoral fraud are negligible or nonexistent. For example, suppose a non-trivial fraction of Democrats (but no Republicans) had the ability to 1) selectively invalidate ballots cast for their opponents *and* 2) perfectly predict what the true vote share would be without interfering with the vote counting process. In this scenario, suppose the Democrats followed the following rule: a) if the “true” vote count would lead to a Republican win, dispute ballots to raise the De-

⁹ With the trivial modification that v_{i2} actually is equal to $Y + e_{i2}$, but e_{i2} has mean zero conditional on V and W , so that $E[v_{i2}|v_{i1}] = E[Y|v_{i1}]$.

mocratic vote share, but b) if the “true” vote count leads to a Democratic win, do nothing. It is easy to see that in repeated elections, this rule would lead to a discontinuous density in v_{i1} right at the $\frac{1}{2}$ threshold.¹⁰

If this kind of fraudulent behavior is important feature of the data, the RDD will lead to invalid inferences; but if it is not, then the RDD is an appropriate design. The important point here is that Proposition 3 (c) implies that the validity of the RDD is empirically testable. That is, if this form of electoral fraud is empirically important, then *all* pre-determined (prior to year 1) characteristics (X) should be different between the two sides of the discontinuity threshold; if it is unimportant, then X should have the same distribution on either side of the threshold.

3.3 Data Issues

Data on U.S. Congressional election returns from 1946-1998 are used in the analysis. In order to use all pairs of consecutive elections for the analysis, the dependent variable v_{i2} is effectively dated from 1948 to 1998, and the independent (score) variable v_{i1} runs from 1946 to 1996. Due to redistricting every 10 years, and since both lags and leads of the vote share will be used, all cases where the independent variable is from a year ending in ‘0’ and ‘2’ are excluded. Because of possible dependence over time, standard errors are clustered at the decade-district level.

In virtually all Congressional elections, the strongest two parties will be the Republicans and the Democrats, but third parties do obtain some small share of the vote. As a result, the cutoff that determines the winner will not be exactly 50 percent. To address this, the main vote share variable is the Democratic vote share minus the vote share of the strongest opponent, which in most cases is a Republican nominee. The Democrat wins the election when this variable “Democratic vote share margin of victory” crosses the 0 threshold, and loses the election otherwise.

Incumbency advantage estimates are reported for the Democratic party only. In a strictly two-party system, estimates for the Republican party would be an exact mirror image, with numerically identical

¹⁰ Note that other “rules” describing fraudulent behavior would nevertheless lead to a continuous density in v_{i1} . For example, suppose all Democrats had the ability to invalidate ballots during the actual vote counting process. Even if this behavior is rampant, if this ability stops when 90 percent of the vote is counted, there is still unpredictability in the vote share tally for the remaining 10 percent of the ballots. It is plausible that the probability density for the vote share in the remaining votes is continuous.

results, since Democratic victories and vote shares would have one-to-one correspondences with Republican losses and vote shares.

The incumbency advantage is analyzed at the level of the party at the district level. That is, the analysis focuses on the advantage to the party from holding the seat, irrespective of the identity of the nominee for the party. Estimation of the analogous effect for the individual candidate is complicated by selective “drop-out”. That is, candidates, whether they win or lose an election, are not compelled to run for (re-)election in the subsequent period. Thus, even a true randomized experiment would be corrupted by this selective attrition.¹¹ Since the goal is to highlight the parallels between RDD and a randomized experiment, to circumvent the candidate drop-out problem, the estimates are constructed at the district level; when a candidate runs uncontested, the opposing party is given a vote share of 0.

Four measures of the success of the party in the subsequent election are used: 1) the probability that the party’s candidate will *both* become the party’s nominee and win the election, 2) the probability that the party’s candidate will become the nominee in the election, 3) the party’s vote share (irrespective of who is the nominee), and 4) the probability that the party wins the seat (irrespective of who is the nominee). The first two outcomes measure the causal impact of a Democratic victory on the political future of the candidate, and the latter two outcomes measure the causal impact of a Democratic victory on the party’s hold on the district seat.

Further details on the construction of the data set is provided in Appendix A.

3.4 RDD Estimates

Figure IIa illustrates the regression discontinuity estimate of the incumbency advantage. It plots the estimated probability of a Democrat both running in and winning election $t + 1$ as a function of the Democratic vote share margin of victory in election t . The horizontal axis measures the Democratic vote share minus the vote share of the Democrats’ strongest opponent (virtually always a Republican). Each point is an aver-

¹¹ An earlier draft (Lee 2000) explores what restrictions on strategic interactions between the candidates can be placed to pin down the incumbency advantage for the candidate for the subpopulation of candidates who would run again whether or not they lose the initial election. A bounding analysis suggests that most of the incumbency advantage may be due to a “quality of candidate” selection effect, whereby the effect on drop-out leads to, on average, weaker nominees for the party in the next election.

age of the indicator variable for running in and winning election $t + 1$ for each interval, which is 0.005 wide. To the left of the dashed vertical line, the Democratic candidate lost election t ; to the right, the Democrat won.

As apparent from the figure, there is a striking discontinuous jump, right at the 0 point. Democrats who barely win an election are much more likely to run for office and succeed in the next election, compared to Democrats who barely lose. The causal effect is enormous: about 0.45 in probability. Nowhere else is a jump apparent, as there is a well-behaved, smooth relationship between the two variables, except at the threshold that determines victory or defeat.

Figures IIIa, IVa, and Va present analogous pictures for the three other electoral outcomes: whether or not the Democrat remains the nominee for the party in election $t + 1$, the vote share for the Democratic party in the district in election $t + 1$, and whether or not the Democratic party wins the seat in election $t + 1$. All figures exhibit significant jumps at the threshold. They imply that for the individual Democratic candidate, the causal effect of winning an election on remaining the party's nominee in the next election is about 0.40 in probability. The incumbency advantage for the Democratic party appears to be about 7 or 8 percent of the vote share. In terms of the probability that the Democratic party wins the seat in the next election, the effect is about 0.35.

In all four figures, there is a positive relationship between the margin of victory and the electoral outcome. For example, as in Figure IVa, the Democratic vote shares in election t and $t + 1$ are positively correlated, both on the left and right side of the figure. This indicates selection bias; a simple comparison of means of Democratic winners and losers would yield biased measures of the incumbency advantage. Note also that Figures IIa, IIIa, and Va exhibit important nonlinearities: a linear regression specification would hence lead to misleading inferences.

Table I presents evidence consistent with the main implication of Proposition 3: in the limit, there is randomized variation in treatment status. The third to eighth rows of Table I are averages of variables that are determined before t , and for elections decided by narrower and narrower margins. For example, in the third row, among the districts where Democrats won in election t , the average vote share for the Democrats

in election $t - 1$ was about 68 percent; about 89 percent of the $t - 1$ elections had been won by Democrats, as the fourth row shows. The fifth and seventh rows report the average number of terms the Democratic candidate served, and the average number of elections in which the individual was a nominee for the party, as of election t . Again, these characteristics are already determined at the time of the election. The sixth and eighth rows report the number of terms and number of elections for the Democratic candidates' strongest opponent. These rows indicate that where Democrats win in election t , the Democrat appears to be a relatively stronger candidate, and the opposing candidate weaker, compared to districts where the Democrat eventually loses election t . For each of these rows, the differences become smaller as one examines closer and closer elections – as c) of Proposition 3 would predict.

These differences persist when the margin of victory is less than 5 percent of the vote. This is, however, to be expected: the sample average in a narrow neighborhood of a margin of victory of 5 percent is in general a biased estimate of the true conditional expectation function at the 0 threshold when that function has a nonzero slope. To address this problem, polynomial approximations are used to generate simple estimates of the discontinuity gap. In particular, the dependent variable is regressed on a fourth-order polynomial in the Democratic vote share margin of victory, separately for each side of the threshold. The final set of columns report the parametric estimates of the expectation function on either side of the discontinuity. Several non-parametric and semi-parametric procedures are also available to estimate the conditional expectation function at 0. For example, Hahn, Todd, and van der Klaauw (2001) suggest local linear regression, and Porter (2003) suggests adapting Robinson's (1988) estimator to the RDD.

The final columns in Table I show that when the parametric approximation is used, all remaining differences between Democratic winners and losers vanish. No differences in the third to eighth rows are statistically significant. These data are consistent with implication c) of Proposition 3, that all pre-determined characteristics are balanced in a neighborhood of the discontinuity threshold. Figures IIb, IIIb, IVb, and Vb, also corroborate this finding. These lower panels examine variables that have already been determined as of election t : the average number of terms the candidate has served in Congress, the average number of times he has been a nominee, as well as electoral outcomes for the party in election $t - 1$. The

figures, which also suggest that the fourth order polynomial approximations are adequate, show a smooth relation between each variable and the Democratic vote share margin at t , as implied by c) of Proposition 3.

The only differences in Table I that do not vanish completely as one examines closer and closer elections, are the variables in the first two rows of Table I. Of course, the Democratic vote share or the probability of a Democratic victory in election $t + 1$ is determined after the election t . Thus the discontinuity gap in the final set of columns represents the RDD estimate of the causal effect of incumbency on those outcomes.

In the analysis of randomized experiments, analysts often include baseline covariates in a regression analysis to reduce sampling variability in the impact estimates. Because the baseline covariates are independent of treatment status, impact estimates are expected to be somewhat insensitive to the inclusion of these covariates. Table II shows this to be true for these data: the results are quite robust to various specifications. Column (1) reports the estimated incumbency effect when the vote share is regressed on the victory (in election t) indicator, the quartic in the margin of victory, and their interactions. The estimate should and does exactly match the differences in the first row of the last set of columns in Table I. Column (2) adds to that regression the Democratic vote share in $t - 1$ and whether they won in $t - 1$. The coefficient on the Democratic share in $t - 1$ is statistically significant. Note that the coefficient on victory in t does not change very much. The coefficient also does not change when the Democrat and opposition political and electoral experience variables are included in Columns (3)-(5).

The estimated effect also remains stable when a completely different method of controlling for pre-determined characteristics is utilized. In Column (6), the Democratic vote share $t + 1$ is regressed on all pre-determined characteristics (variables in rows three through eight), and the discontinuity jump is estimated using the *residuals* of this initial regression as the outcome variable. The estimated incumbency advantage remains at about 8 percent of the vote share. This should be expected if treatment is locally independent of all pre-determined characteristics. Since the average of those variables are smooth through the threshold, so should be a linear function of those variables. This principle is demonstrated in Column

(7), where the vote share in $t - 1$ is subtracted from the vote share in $t + 1$ and the discontinuity jump in that difference is examined. Again, the coefficient remains at about 8 percent.

Column (8) reports a final specification check of the regression discontinuity design and estimation procedure. I attempt to estimate the “causal effect” of winning in election t on the vote share in $t - 1$. Since we *know* that the outcome of election t cannot possibly causally effect the electoral vote share in $t - 1$, the estimated impact should be zero. If it significantly departs from zero, this calls into question, *some* aspect of the identification strategy and/or estimation procedure. The estimated effect is essentially 0, with a fairly small estimated standard error of 0.011. All specifications in Table II were repeated for the indicator variable for a Democrat victory in $t + 1$ as the dependent variable, and the estimated coefficient was stable across specifications at about 0.38 and it passed the specification check of Column (8) with a coefficient of -0.005 with a standard error of 0.033.

In summary, the econometric model of election returns outlined in the previous section allows for a great deal of non-random selection. The seemingly mild continuity assumption on the distribution of v_{i1} results in the strong prediction of local independence of treatment status (Democratic victory) that itself has an “infinite” number of testable predictions. The distribution of *any* variable determined prior to assignment must be virtually identical on either side of the discontinuity threshold. The empirical evidence is consistent with these predictions, suggesting that even though U.S. House elections are non-random selection mechanisms – where outcomes are influenced by political actors – they also contain randomized experiments that can be exploited by RD analysis.¹²

3.5 Comparison to Existing estimates of the Incumbency Advantage

It is difficult to make a direct comparison between the above RDD estimates and existing estimates of the incumbency advantage in the political science literature. This is because the RDD estimates identify a different, but related concept. The existing literature generally focuses on the concept of an incumbent *legislator* advantage, while the RDD approach identifies an overall incumbent *party* advantage.

¹² This notion of using “as good as randomized” variation in treatment from close elections has been utilized in Miguel and Zaidi (2003), Clark (2004), Linden (2004), Lee, Moretti, and Butler (2004), DiNardo and Lee (2004).

Measuring the incumbent legislator advantage answers the following question: From the party's perspective, what is the electoral gain to having the incumbent *legislator* run for re-election, relative to having a new candidate *of the same party* run in the same district?¹³ This incumbency advantage is the electoral success that an incumbent party enjoys if the incumbent runs for re-election, *over and above* the electoral outcome that *would have* occurred if a new nominee for the party had run in the same district.¹⁴

By contrast, measuring the incumbent party advantage answers the following question: From the party's perspective, what is the electoral gain to being the incumbent party in a district, relative to *not* being the incumbent party? In other words, while the incumbent party's vote share typically exceeds 50 percent, some of the votes would have been gained by the party even if it had *not* been in control of the seat.

The electoral advantage to being the incumbent party potentially works through a number of different mechanisms, and *includes* as a possible mechanism the incumbent legislator advantage. That is, when the Democratic nominee barely wins an election, it raises the probability that the nominee will run for re-election as an incumbent legislator. Insofar as there are advantages to being an incumbent legislator as opposed to a new nominee, this will contribute to the overall incumbent party advantage identified by the RDD. In addition, the incumbent party advantage includes the gain that would have occurred even if the victorious candidate in the first election did *not* run for re-election.

The interested reader is referred to Appendix B, which explains the difference between the RDD estimate of the incumbency advantage and estimates typically found in the political science literature.

4 Conclusion

In one sense, the RDD is no different from all other research designs: causal inferences are only

¹³ The most precise statement of the counterfactual can be found in Gelman and King [1990], who use "potential outcomes" notation, to define the "incumbency advantage". They define the incumbency advantage in a district as a the difference between "the proportion of the vote received by the incumbent legislator in his or her district..." and the "proportion of the vote received by the incumbent party in that district, if the incumbent legislator does not run..."

¹⁴ This notion is also expressed in Alford and Brady [1993], who note that incumbents can incidentally benefit from the party holding the seat, and that the personal incumbency advantage should be differentiated from the party advantage, and that "[i]t is this concept of personal incumbency advantage that most of the incumbency literature, and the related work in the congressional literature, implicitly turns on." Examples of typical incumbency studies that utilize this concept of incumbency include: Payne [1980], Alford and Hibbing [1981], Collie [1981], and Garand and Gross [1984]. More recently, work by Ansolabehere and Snyder [2001, 2002], and Ansolabehere, Snyder, and Stewart [2000] implicitly examine this concept by examining the coefficient on the same incumbency variable defined in Gelman and King [1990]: 1 if a Democrat incumbent seeks re-election, 0 if it is an open seat, and -1 if a Republican seeks re-election.

possible as a direct result of key statistical assumptions about how treatment is assigned to individuals. The key assumption examined in this paper is the continuity of the density of V for each individual. What makes the treatment assignment mechanism described in this paper somewhat distinctive is that what appears to be a weak continuity assumption directly leads to very restrictive statistical properties for treatment status. Randomized variation in treatment – independence (local to the threshold) – is perhaps the most restrictive property possible, with potentially an infinite number of testable restrictions, corresponding to the number of pre-determined characteristics and the number of available moments for each variable. These testable restrictions are an advantage when seeking to subject the research design to a battery of over-identifying tests.¹⁵

Although the continuity assumption appears to be a weak restriction – particularly since it is implicitly made in most selection models in econometrics – there are reasons to believe they might be violated when agents have direct and exact control over the score V . If there are benefits to receiving the treatment, it is natural to expect those who gain the most to choose their value of V to be above – and potentially just marginally above – the relevant threshold. Ruling out this kind of behavior appears to be an important part of theoretically justifying the application of the RDD in any particular context.

¹⁵ In addition to testing whether baseline characteristics are similar in the marginal treated and control groups, one can examine if there is a discontinuity in the marginal distribution of V . See McCrary [2004]

Appendix A. Description of Data

The data used for this analysis is based on the candidate-level Congressional election returns for the U.S., from ICPSR study 7757, “Candidate and Constituency Statistics of Elections in the United States, 1788-1990”.

The data were initially checked for internal consistencies (e.g. candidates’ vote totals not equalling reported total vote cast), and corrected using published and official sources (Congressional Quarterly [1997] and the United States House of Representatives Office of the Clerk’s Web Page). Election returns from 1992-1998 were taken from the United States House of Representatives Office of the Clerk’s Web Page, and appended to these data. Various states (e.g. Arkansas, Louisiana, Florida, and Oklahoma) have laws that do not require the reporting of candidate vote totals if the candidate ran unopposed. If they are the only candidate in the district, they were assigned a vote share of 1. Other individual missing vote totals were replaced with valid totals from published and official sources. Individuals with more than one observation in a district year (e.g. separate Liberal and Democrat vote totals for the same person in New York and Connecticut) were given the total of the votes, and were assigned to the party that gave the candidate the most votes. The name of the candidate was parsed into last name, first name, and middle names, and suffixes such as “Jr., Sr., II, III, etc.”

Since the exact spelling of the name differs across years, the following algorithm was used to create a unique identifier for an individual that could match the person over time. Individuals were first matched on state, first 5 characters of the last name, and first initial of the first name. The second layer of the matching process isolates those with a suffix such as Jr. or Sr., and small number of cases were hand-modified using published and official sources. This algorithm was checked by drawing a random sample of 100 election-year-candidate observations from the original sample, tracking down every separate election the individual ran in (using published and official sources; this expanded the random sample to 517 election-year-candidate observations), and asking how well the automatic algorithm performed. The fraction of observations from this “truth” sample that matched with the processed data was 0.982. The

fraction of the processed data for which there was a “true” match was 0.992. Many different algorithms were tried, but the algorithm above performed best based on the random sample.

Throughout the sample period (1946-1998), in about 3 percent of the total possible number of elections (based on the number of seats in the House in each year), no candidate was reported for the election. I impute the missing values using the following algorithm. Assign the state-year average electoral outcome; if still missing, assign the state-decade average electoral outcome.

Two main data sets are constructed for the analysis. For all analysis at the Congressional level, I keep all years that do not end in ‘0’ or ‘2’. This is because, strictly speaking, Congressional districts cannot be matched between those years, due to decennial redistricting, and so in those years, the previous or next electoral outcome is undefined. The final data set has 6558 observations. For the analysis at the individual candidate level, one can use more years, because, despite redistricting, it is still possible to know if a candidate ran in *some* election, as well as the outcome. This larger dataset has 9674 Democrat observations.

For the sake of conciseness, the empirical analysis in the paper focuses on observations for Democrats only. This is done to avoid the “double-counting” of observations, since in a largely two-party context, a winning Democrat will, by construction, produce a losing Republican in that district and vice versa. (It is unattractive to compare a close winner to the closer loser in the same district) In reality, there are third-party candidates, so a parallel analysis done by focusing on Republican candidates will not give a literal mirror image of the results. However, since third-party candidates tend not to be important in the U.S. context, it turns out that all of the results are qualitatively the same, and are available from the author upon request.

Appendix B. Comparison of Regression Discontinuity Design and other estimates of the Incumbency Advantage

Reviews of the existing methodological literature in Gelman and King [1990], Alford and Brady [1993], and Jacobson [1997] suggest that most of the research on incumbency are variants of three general approaches – the “sophomore surge”, the “retirement slump”, and the Gelman-King index [1990].¹⁶ Appendix Figures Ia and Ib illustrate the differences between these three approaches and the regression discontinuity strategy employed in this paper. They also clearly show that the three commonly-used approaches to measuring the “incumbency advantage” are estimating an *incumbent legislator advantage* as opposed to an *incumbent party advantage*. The regression discontinuity design is ideally suited for estimating the latter incumbency advantage.

Appendix Figure Ia illustrates the idea behind the “sophomore surge”. The solid line shows a hypothetical relationship between the average two-party Democratic vote share in period 2 – V_2 – as a function of the Democratic vote share in period 1, V_1 . In addition, the dotted line just below the solid line on the right side of the graph represents the average V_2 as a function of V_1 , for the sub-sample of elections that were won by first-time Democrats in period 1.¹⁷ The idea behind the sophomore surge is to subtract from the average V_2 for all Democratic first-time incumbents (\bar{V}_2), an amount that represents the strength of the party in those districts apart from any incumbent legislator advantage. The “sophomore surge” approach subtracts off \bar{V}_1 , the average V_1 for those same Democratic first-time incumbents.

Appendix Figure Ib illustrates how the “retirement slump” is a parallel measure to the “sophomore surge”. In this figure, the dashed line below the solid line represents the average V_2 as a function of V_1 , for those districts that will have open seats as of period 2. In other words, it is the relationship for those districts in which the Democratic incumbent retires and does not seek re-election in period 2. Here, the idea is to subtract from the average vote share gained by the retiring Democratic incumbents (\bar{V}_1), an amount

¹⁶ See Gelman and King [1990] for a concise review of existing methods as of 1990. More recently, Levitt and Wolfram [1997] use a “modified sophomore surge” approach and Cox and Katz [1996] use a specification “adapted from Gelman and King [1990]”.

¹⁷ In principle, a “mirror-image” line exists for the Republican first-time incumbents. But I omit the line to make the graph clearer. Also, note that in this graph, to simplify exposition, I am assuming that all first-time incumbents seek re-election. The basic ideas hold when relaxing this assumption, but the notation is slightly more cumbersome.

that reflects the strength of the party. The “retirement slump” approach subtracts off \bar{V}_2 , the average V_2 for the incoming Democratic candidate in those districts.

Appendix Figure Ib also illustrates Gelman and King’s approach to measuring incumbency advantage. The dotted line above the solid line is the average V_2 , as a function of V_1 , for those districts in which the Democratic incumbent is seeking re-election. The idea behind their approach is to subtract from the average vote share V_2 gained by incumbents seeking re-election, an amount that reflects how the party would perform if a new candidate ran for the party, *while controlling for the lagged vote share* V_1 . Thus, the gap between the parallel dashed and dotted lines in Appendix Figure Ib represent the incumbent legislator advantage measured by the Gelman-King index (ψ).¹⁸

Finally, Appendix Figures Ia and Ib illustrate that the approach of the regression discontinuity design isolates a different aspect of the incumbency advantage – the incumbent party advantage. The idea is to make a comparison of the electoral performance in period 2 of the Democratic party between districts that were *barely won* (say, by 0.1 percent of the vote) and districts that were *barely lost* by the Democratic party in period 1. Thus the regression discontinuity estimate (β) is depicted by the discontinuous jumps in the solid lines at the 0.50 threshold in Appendix Figures Ia and Ib.¹⁹

The figures show that the discontinuity gap directly addresses the counterfactual question that is at the heart of measuring the incumbent party advantage: *How would the Democratic party have performed in period 2, had they not held the seat (i.e. had the Democrats lost the election in period 1)?* The best way to estimate that quantity is to examine elections where the Democrats just barely missed being the incumbent party for the election in period 2 – the point on the solid line just to the left of the 0.50 threshold.

It is important to recognize that the incumbent party advantage includes in part the incumbent legislator advantage, since a Democrat close win, for example, raises the probability that the Democratic nominee runs for re-election as an incumbent legislator in the next election. It also includes the vote share gain (or loss) that occurs because the party *per se* holds the seat.

¹⁸ Again, the mirror-image dotted and dashed lines for Republicans are ignored to make the exposition clearer in the graph.

¹⁹ While it is tempting to equate β here to β_2 in equation 6 of Gelman and King [1990], this would be incorrect, because Gelman and King include their key variable, I_2 , in the specification. (If they did not include I_2 , their key variable, then β_2 would equal β here). Thus, the regression discontinuity estimates presented in this paper cannot be recovered from a Gelman-King-type analysis.

Nevertheless, it is also important to recognize that in principle, there is no necessary connection between the RD estimates and the other estimates. It is possible that the regression discontinuity estimate could be zero (no break in the solid lines at 0.50), while at the same time the sophomore surge, retirement slump, and Gelman-King index could be significant. Alternatively, in principle, there could be a large estimate of β , while at the same time the other measures could be zero.

A final important point is that the sophomore surge relies upon situations where either an incumbent is defeated or a seat is thrown open, and the retirement slump, and Gelman-King index relies upon situations where seats are thrown open by an incumbent that does not seek re-election. In principle, for any year in which every incumbent becomes the nominee in a re-election bid (and ignoring redistricting years) it would be *impossible* to estimate any of those three measures; the dotted and dashed lines in Appendix Figures Ia and Ib would not exist. In this case, the only incumbency advantage concept that *could* be measured would be the incumbent party advantage identified by the RDD.

Appendix C. Proofs

Condition A1. Random draw from population. Let μ be a probability measure on (Ω, \mathcal{F}) . Each $\omega \in \Omega$ represents an individual. $(\Omega, \mathcal{F}, \mu)$ describes the probabilities of drawing individuals from a (possibly infinite) population.

Condition A2. Stochastic treatment assignment. For each $\omega \in \Omega$, let v_ω be a probability measure on (Δ, \mathcal{D}) . $(\Delta, \mathcal{D}, v_\omega)$ describes the probabilities associated with receiving the treatment (or, in the RDD, the score V), for each individual ω . Assume that for any $B \in \mathcal{D}$, $v_\omega(B)$ as a function of ω is measurable \mathcal{F} . Let \mathcal{G} be the σ -field consisting of all sets $\Omega \times A$, where $A \in \mathcal{D}$.

Condition A3. Probabilities for the overall experiment. Define P as follows: $\forall E \in \mathcal{F} \times \mathcal{D}$, $P(E) = \int_{\Omega} v_\omega[\delta : (\omega, \delta) \in E] \mu(d\omega)$. It can be shown that P is a probability measure on $(\Omega \times \{0, 1\}, \mathcal{F} \times \mathcal{D})$.

Condition A4. Pre-determined characteristics. Let $X = x(\omega)$ be a real-valued function that is measurable $\mathcal{F} \times \mathcal{D}$. It follows that it is also measurable \mathcal{F} .

Condition A5. Finite first moments. E_P and E_μ denote expectations with respect to probability measures P and μ , respectively. Where appropriate, Y , Y_1 , Y_0 , $\frac{f_\omega(0)}{f(0)}Y$, $\frac{f_\omega(0)}{f(0)}Y_1$, and $\frac{f_\omega(0)}{f(0)}Y_0$ are each assumed to be integrable P and integrable μ .

Condition B1. Binary treatment assignment model. Let $\Delta = \{0, 1\}$ and $\mathcal{D} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. Define the random variable D as $D = \delta$, $\delta \in \Delta$, which is measurable $\mathcal{F} \times \mathcal{D}$.

Condition B2. Regression discontinuity design. Let $\Delta = \mathbb{R}$, and $\mathcal{D} = \mathcal{R}^1$ be the class of linear Borel sets. Define the random variable V – measurable on $\mathcal{F} \times \mathcal{D}$ – as $V(\delta) = \delta$, $\delta \in \Delta$, and let $D = 1[V \geq 0]$.

Condition C1. Potential outcomes. Let $Y_1 = y_1(\omega)$, $Y_0 = y_0(\omega)$, be real-valued functions that are measurable $\mathcal{F} \times \mathcal{D}$ (and hence measurable \mathcal{F}). Let $Y = DY_1 + (1 - D)Y_0$.

Condition C2. Potential outcome function. Let $Y = y(\omega, \delta)$ be a real-valued function that is measurable $\mathcal{F} \times \mathcal{D}$. Let $y(\cdot, \cdot)$ be continuous in the second argument except at $\delta = 0$, where the function is only continuous from the right. Define the function $Y^+ = y(\omega, 0)$ and $Y^- = \lim_{\epsilon \rightarrow 0^+} y(\omega, -\epsilon)$.

Condition D1. Treatment randomization. v_ω is identical for all $\omega \in \Omega$

Condition D2. Continuous density of score. Let $F_\omega(\delta) = v_\omega(-\infty, \delta]$, and $f_\omega(\delta)$ its derivative with respect to δ . Let $f(\delta) = \int_\Omega f_\omega(\delta) \mu(d\omega)$. Assume that $0 < f_\omega(\delta)$, and $f_\omega(\delta)$ is continuous in δ on \mathbb{R} . (Note that if v_ω is measurable \mathcal{F} , one can show that in this set-up, so too are F_ω and f_ω).

Proposition 1. If Conditions A1-A5, B1, C1, and D1 hold, then:

- a) $\forall F \in \mathcal{F}, P[F \times \Delta | D = 1] = P[F \times \Delta | D = 0] = P[F \times \Delta] = \mu[F]$
- b) $E_P[Y | D = 1] - E_P[Y | D = 0] = E_\mu[Y_1 - Y_0] \equiv ATE$
- c) $\forall x_0 \in \mathbb{R}, P[X \leq x_0 | D = 1] = P[X \leq x_0 | D = 0] = P[X \leq x_0] = \mu[\omega : X \leq x_0]$

Proof. a) $P[F \times \Delta | D = 1] = P[(F \times \Delta) \cap (\Omega \times \{1\})] / P[\Omega \times \{1\}]$. Numerator is $\int_{F \times \{1\}} P(d(\omega, \delta))$.

This is equal to $\int_F \left[\int_{\{1\}} v_\omega(d\delta) \right] \mu(d\omega) = v_\omega(\{1\}) \cdot \mu[F]$ by 18.20.c of Billingsley (1995) and by D1.

Similarly, denominator is $v_\omega(\{1\})$. Similar argument holds for $P[F \times \Delta | D = 0]$. b) Need to show that

conditional expectation of Y_1 given \mathcal{G} , evaluated at $D = 1$ is equal to $E_\mu[Y_1]$. It can be shown that

the conditional expectation of Y given \mathcal{G} can be written as $\alpha(\delta_0) \equiv \frac{1}{P[\Omega \times \{\delta_0\}]} \int_{\Omega \times \{\delta_0\}} Y_{\delta_0} P(d(\omega, \delta))$,

for $\delta_0 = 0$ and 1 . Consider the case when $\delta_0 = 1$. We then have $\frac{1}{P[\Omega \times \{1\}]} \int_{\Omega \times \{1\}} Y_1 P(d(\omega, \delta)) =$

$\frac{1}{P[\Omega \times \{1\}]} \int_\Omega \left[\int_{\{1\}} Y_1 v_\omega(d\delta) \right] \mu(d\omega)$ by 18.20.c of Billingsley (1995). Because Y_1 is only a function of ω ,

and by D1, this becomes $\frac{v_\omega(\{1\})}{P[\Omega \times \{1\}]} \int_\Omega Y_1 \mu(d\omega)$ which is equal to $\int_\Omega Y_1 \mu(d\omega) = E_\mu[Y_1]$; a similar argu-

ment shows that $\alpha(0) = E_\mu[Y_0]$. c) By A4, for every $x_0 \in \mathbb{R}$, $F \equiv [\omega : X(\omega) \leq x_0]$ is in \mathcal{F} , and thus c)

follows from a).

Proposition 2 If Conditions A1-A5, B2, C1, and D2 hold, then:

- a) $\forall F \in \mathcal{F}, P[F \times \Delta | V = v]$ is continuous in v at $v = 0$
- b) $E_P[Y | V = 0] - \lim_{\varepsilon \rightarrow 0^+} E_P[Y | V = -\varepsilon] = E_P[Y_1 - Y_0 | V = 0] = E_\mu \left[\frac{f_\omega(0)}{f(0)} (Y_1 - Y_0) \right] \equiv ATE^*$
- c) $\forall x_0 \in \mathbb{R}, P[X \leq x_0 | V = v]$ is continuous in v at $v = 0$

Proof. a) Fix $F \in \mathcal{F}$, and consider the function $\alpha : \Omega \times \Delta \rightarrow \mathbb{R}, \alpha(z, \delta) \equiv \frac{\int_F f_\omega(\delta) \mu(d\omega)}{f(\delta)}$. It

suffices to show 1) that $\alpha(z, \delta)$ is a version of the conditional probability of $F \times \Delta$ given \mathcal{G} , and 2) that

$\alpha(z, \delta)$ is continuous in δ on \mathbb{R} . First, for each $\Omega \times A$ we have – by 18.20.c and 18.20.d of Billingsley

(1995) $-\int_{\Omega \times A} \alpha(z, \delta) P(d(z, \delta)) = \int_A \frac{\int_F f_\omega(\delta) \mu(d\omega)}{f(\delta)} v(d\delta)$, where v is a probability measure defined by $v(B) = \int_\Omega v_\omega(B) \mu(d\omega)$, for all $B \in \mathcal{D}$. v has density f with respect to Lebesgue measure because for all $B \in \mathcal{D}$, $\int_B f(\delta) d\delta = \int_B [\int_\Omega f_\omega(\delta) \mu(d\omega)] d\delta = \int_\Omega [\int_B f_\omega(\delta) d\delta] \mu(d\omega) = \int_\Omega v_\omega(B) \mu(d\omega)$, by Fubini's theorem, and because $f_\omega(\delta)$ is a density of v_ω . Thus, by theorem 16.11 of Billingsley (1995), $\int_A \frac{\int_F f_\omega(\delta) \mu(d\omega)}{f(\delta)} v(d\delta) = \int_A [\int_F f_\omega(\delta) \mu(d\omega)] d\delta$, which equals $\int_F [\int_A f_\omega(\delta) d\delta] \mu(d\omega)$, by Fubini's theorem. This equals $\int_F v_\omega(A) \mu(d\omega) = P[F \times A]$, because f_ω is a density and by 18.20.c of Billingsley (1995).

Second, to show continuity of $\alpha(z, \delta)$, it suffices to show that for any $F \in \mathcal{F}$ and any sequence $\delta_n \rightarrow 0$, $\int_F f_\omega(\delta_n) \mu(d\omega) \rightarrow \int_F f_\omega(0) \mu(d\omega)$. This follows from dominated convergence, noting that $f_\omega(\delta_n) \leq g_\omega$, if $g_\omega \equiv \sup_n f_\omega(\delta_n)$, which is finite for each ω , because $f_\omega(\delta_n)$ converges to $f_\omega(0)$, by D2.

b) Consider the function $\beta : \Omega \times \Delta \rightarrow \mathbb{R}$, $\beta(z, \delta) = \int_\Omega Y \frac{f_\omega(\delta)}{f(\delta)} \mu(d\omega)$. It suffices to show that 1) $\beta(z, \delta)$ is a version of the conditional expectation of Y given \mathcal{G} , and 2) $\beta(z, 0) = E_P[Y_1|V=0] = E_\mu\left[\frac{f_\omega(0)}{f(0)} Y_1\right]$ and $\lim_{\varepsilon \rightarrow 0^+} \beta(z, -\varepsilon) = E_P[Y_0|V=0] = E_\mu\left[\frac{f_\omega(0)}{f(0)} Y_0\right]$. First, for all $\Omega \times A \in \mathcal{G}$, we have $\int_{\Omega \times A} \beta(z, \delta) P(d(z, \delta)) = \int_A [\int_\Omega Y \frac{f_\omega(\delta)}{f(\delta)} \mu(d\omega)] v(d\delta)$ by 18.20.c and 18.20.d of Billingsley (1995). This is equal to $\int_\Omega [\int_A Y \frac{f_\omega(\delta)}{f(\delta)} v(d\delta)] \mu(d\omega) = \int_\Omega [\int_A Y f_\omega(\delta) d\delta] \mu(d\omega)$ because v has density f (see above). This is equal to $\int_\Omega [\int_A Y v_\omega(d\delta)] \mu(d\omega) = \int_{\Omega \times A} Y P(d(\omega, \delta))$, because v_ω has density f_ω , and by 18.20.c of Billingsley (1995). Second, let $\delta = 0$. $\int_\Omega Y \frac{f_\omega(0)}{f(0)} \mu(d\omega) = \int_\Omega Y_1 \frac{f_\omega(0)}{f(0)} \mu(d\omega) = E_P[Y_1|V=0]$, by the definition of Y , and the same argument above. Also, $\int_\Omega Y_1 \frac{f_\omega(0)}{f(0)} \mu(d\omega) = E_\mu\left[\frac{f_\omega(0)}{f(0)} Y_1\right]$. Finally, let $\delta_n < 0$, $\delta_n \rightarrow 0$. $\frac{f_\omega(\delta_n)}{f(\delta_n)} \rightarrow \frac{f_\omega(0)}{f(0)}$, by D2. Need to show $\lim_n \int_\Omega Y_0 \frac{f_\omega(\delta_n)}{f(\delta_n)} \mu(d\omega) = \int_\Omega Y_0 \frac{f_\omega(0)}{f(0)} \mu(d\omega)$. This follows from dominated convergence with $|Y_0 \frac{f_\omega(\delta_n)}{f(\delta_n)}|$ dominated by $|Y_0 \frac{g_\omega}{\inf_n f(\delta_n)}|$ (same g_ω as above). By the same argument as above, $\int_\Omega Y_0 \frac{f_\omega(0)}{f(0)} \mu(d\omega) = E_P[Y_0|V=0] = E_\mu\left[\frac{f_\omega(0)}{f(0)} Y_0\right]$.

c) By A4, for every $x_0 \in \mathbb{R}$, $F \equiv [\omega : X(\omega) \leq x_0]$ is in \mathcal{F} , and thus c) follows from a).

Proposition 3

If Conditions A1-A5, B2, C2, and D2 hold, then:

a) and c) of Proposition 2 are true, and

$$\text{b) } E_P [Y|V = 0] - \lim_{\varepsilon \rightarrow 0^+} E_P [Y|V = -\varepsilon] = E_\mu \left[\frac{f_\omega(0)}{f(0)} (Y^+ - Y^-) \right] \equiv AT E^{**}$$

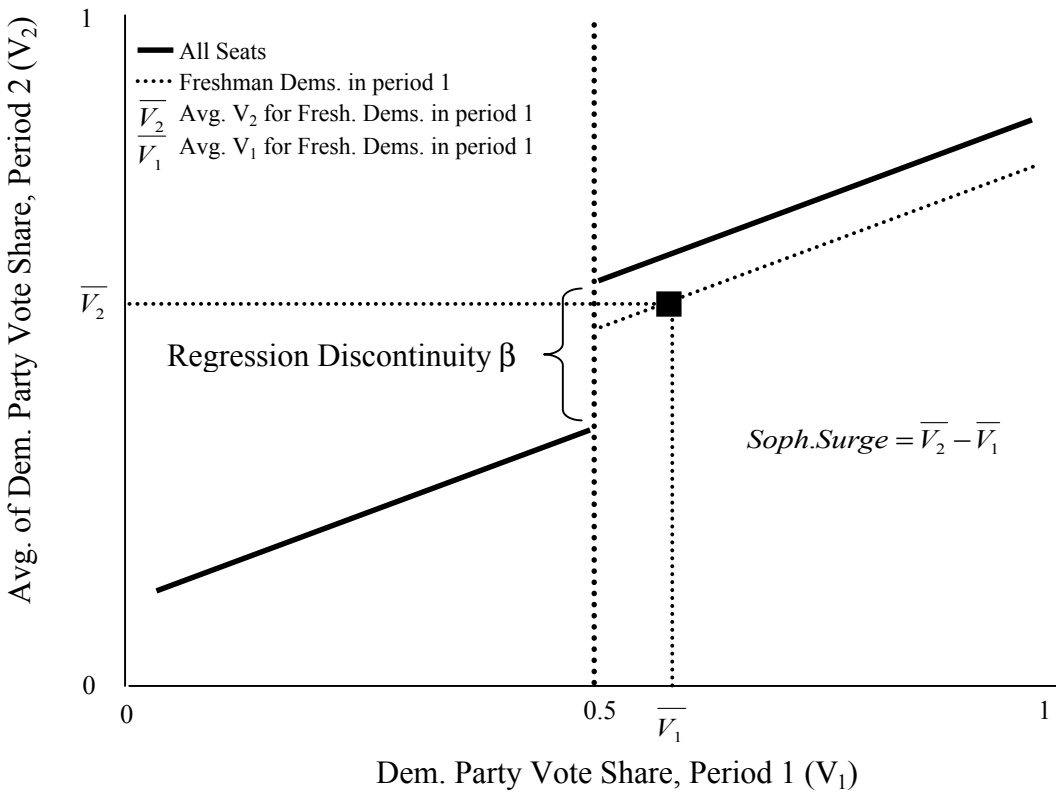
Proof. For a) and c), see the proof to Proposition 2. b) First, following the argument the proof to Proposition 2, $\beta(z, \delta)$ is a version of the conditional expectation of Y given \mathcal{G} . Second, let $\delta = 0$. $\int_\Omega Y \frac{f_\omega(0)}{f(0)} \mu(d\omega) = \int_\Omega Y^+ \frac{f_\omega(0)}{f(0)} \mu(d\omega) = E_\mu \left[\frac{f_\omega(0)}{f(0)} Y^+ \right]$. Finally, let $\delta_n < 0$, $\delta_n \rightarrow 0$. $\frac{f_\omega(\delta_n)}{f(\delta_n)} \rightarrow \frac{f_\omega(0)}{f(0)}$, by D2. Need to show $\lim_n \int_\Omega Y \frac{f_\omega(\delta_n)}{f(\delta_n)} \mu(d\omega) = \int_\Omega Y^- \frac{f_\omega(0)}{f(0)} \mu(d\omega)$. This follows from dominated convergence with $|Y \frac{f_\omega(\delta_n)}{f(\delta_n)}|$ dominated by $|h_\omega \frac{g_\omega}{\inf_n f(\delta_n)}|$ (same g_ω as above) where $h_\omega \equiv \sup_n |y(\omega, \delta_n)|$, which is finite for each ω , because $y(\omega, \delta_n) \rightarrow Y^-$, by C2. It follows that $\int_\Omega Y^- \frac{f_\omega(0)}{f(0)} \mu(d\omega) = E_\mu \left[\frac{f_\omega(0)}{f(0)} Y^- \right]$.

References

- [1] Alford, John R., and David W. Brady. "Personal and Partisan Advantage in U.S. Congressional Elections." In *Congress Reconsidered*, 5th ed., ed. Lawrence C. Dodd and Bruce I. Oppenheimer. Washington, DC: CQ Press, 1993.
- [2] Alford, John R., and John R. Hibbing. "Increased Incumbency Advantage in the House." *Journal of Politics* 43 (1981): 1042-61.
- [3] Angrist, Joshua D., and Victor Lavy. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114 (1998):533-75.
- [4] Ansolabehere, Stephen, and James M. Snyder, "The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Offices, 1942-2000" MIT manuscript, June 2001.
- [5] Ansolabehere, Stephen, and James M. Snyder, "Using Term Limits to Estimate Incumbency Advantages When Officeholders Retire Strategically." MIT manuscript, January 2002.
- [6] Ansolabehere, Stephen, James M. Snyder, and Charles Stewart, "Old Voters, New Voters, and the Personal Vote: Using Redistricting to Measure the Incumbency Advantage." *American Journal of Political Science* 44 (2000):17-34.
- [7] Billingsley, Patrick. *Probability and Measure, Third Edition*. John Wiley & Sons. New York: 1995.
- [8] Clark, Damon. "Politics, Markets and Schools: Quasi-Experimental Evidence on the Impact of Autonomy and Competition from a Truly Revolutionary UK Reform" Manuscript, November, 2004.
- [9] Collie, Melissa P. "Incumbency, Electoral Safety, and Turnover in the House of Representatives, 1952-1976." *American Political Science Review* 75 (1981): 119-31.
- [10] Cox, Gary, and Jonathan Katz. "Why Did the Incumbency Advantage Grow." *American Journal of Political Science* 40 (1996):478-497.
- [11] Congressional Quarterly. *Congressional Elections: 1946-1996*. 1997.
- [12] DiNardo, John, and David S. Lee, "Economic Impacts of New Unionization on Private Sector Employers: 1984-2001." *Quarterly Journal of Economics* 119 (2004): 1383-1442.
- [13] Erikson, Robert S. "The Advantage of Incumbency in Congressional Elections." *Polity* 3 (1971): 395-405.
- [14] Garand, James C., and Donald A. Gross. "Change in the Vote Margins for Congressional Candidates: A Specification of the Historical Trends." *American Political Science Review* 78 (1984): 17-30.
- [15] Gelman, Andrew, and Gary King. "Estimating Incumbency Advantage without Bias." *American Journal of Political Science* 34 (1990): 1142-64.
- [16] Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (2001): 201-209.
- [17] Heckman, J., "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica*, Vol. 46, No. 4, (Jul. 1978), pp. 931-959.
- [18] Inter-university Consortium for Political and Social Research. "Candidate and Constituency Statistics of Elections in the United States, 1788-1990" *Computer File* 5th ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research, producer and distributor, 1995.
- [19] Jacobson, Gary C. *The Politics of Congressional Elections*, Menlo Park, California: Longman, 1997.
- [20] Lee, David S. "Is there Really an Electoral Advantage to Incumbency? Evidence from Close Elections to the United States House of Representatives" *Harvard University Manuscript*, April 2000.
- [21] Lee, David S., Enrico Moretti, and Matthew J. Butler, "Do Voters Affect or Elect Policies? Evidence from the U.S. House." *Quarterly Journal of Economics* 119 (2004): 807-860.
- [22] Levitt, Steven D., and Catherine D. Wolfram. "Decomposing the Sources of Incumbency Advantage in the U.S. House." *Legislative Studies Quarterly* 22 (1997) 45-60.

- [23] Linden, Leigh, "Are Incumbents Really Advantaged? The Preference for Non-Incumbents in Indian National Elections" Columbia University Manuscript, January, 2004.
- [24] Martorell, Francisco, "Do Graduation Exams Matter? A Regression-Discontinuity Analysis of the Impact of Failing the Exit Exam on High School and Post-High School Outcomes" UC Berkeley manuscript, September, 2004.
- [25] Matsudaira, Jordan D., "Sinking or Swimming? Evaluating the Impact of English Immersion vs. Bilingual Education on Student Achievement" University of Michigan manuscript, October, 2004.
- [26] Miguel, Edward, and Farhan Zaidi, "Do Politicians Reward their Supporters? Public Spending and Incumbency Advantage in Ghana," UC Berkeley manuscript, 2003.
- [27] McCrary, Justin, "Testing for Manipulation of the Running Variable in the Regression Discontinuity Design", University of Michigan manuscript, 2004.
- [28] Payne, James L. "The Personal Electoral Advantage of House Incumbents." *American Politics Quarterly* 8 (1980): 375-98.
- [29] Porter, Jack, "Estimation in the Regression Discontinuity Model", *Harvard University manuscript*, May 2003.
- [30] Robinson, P. (1988) "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.
- [31] Thistlethwaite, D., and D. Campbell. "Regression-Discontinuity Analysis: An alternative to the ex post facto experiment." *Journal of Educational Psychology* 51 (1960): 309-17.
- [32] van der Klaauw, Wilbert. "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach," *International Economic Review*, Vol 43(4), November 2002.

**Appendix Figure 1a: Identification of Incumbency Advantage:
Sophomore Surge vs. Regression Discontinuity**



**Appendix Figure 1b: Identification of Incumbency Advantage:
Gelman-King and Retirement Slump vs. Regression Discontinuity**

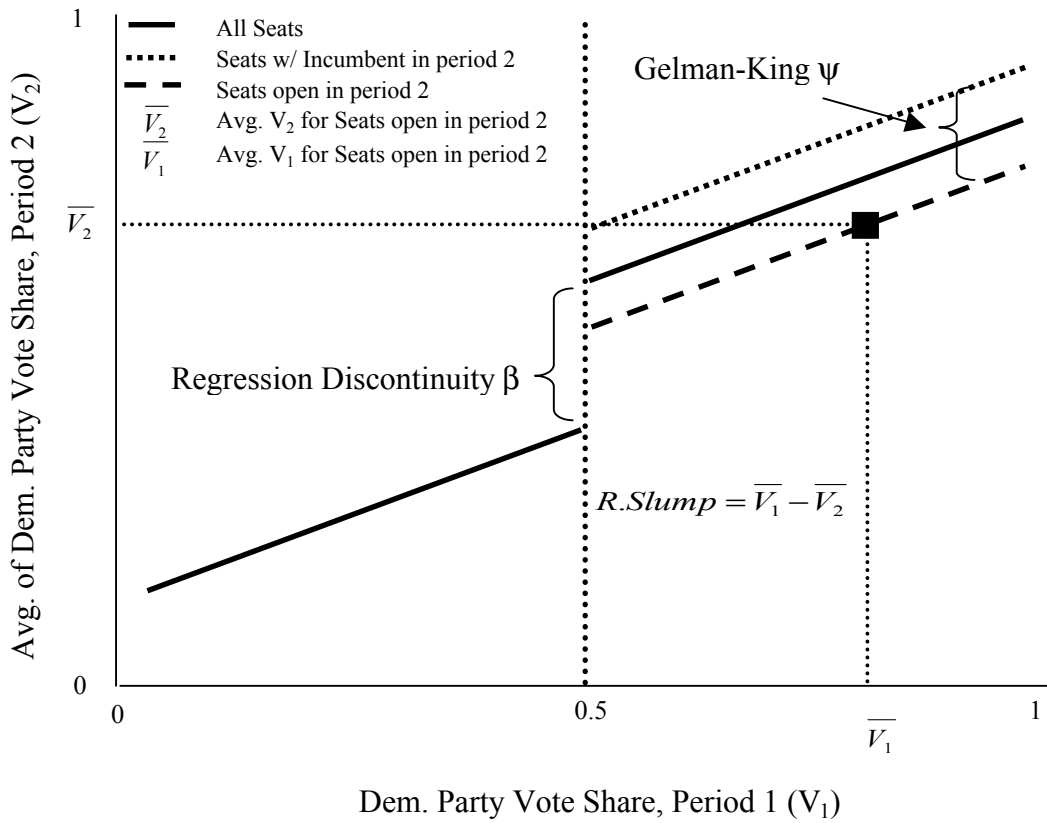
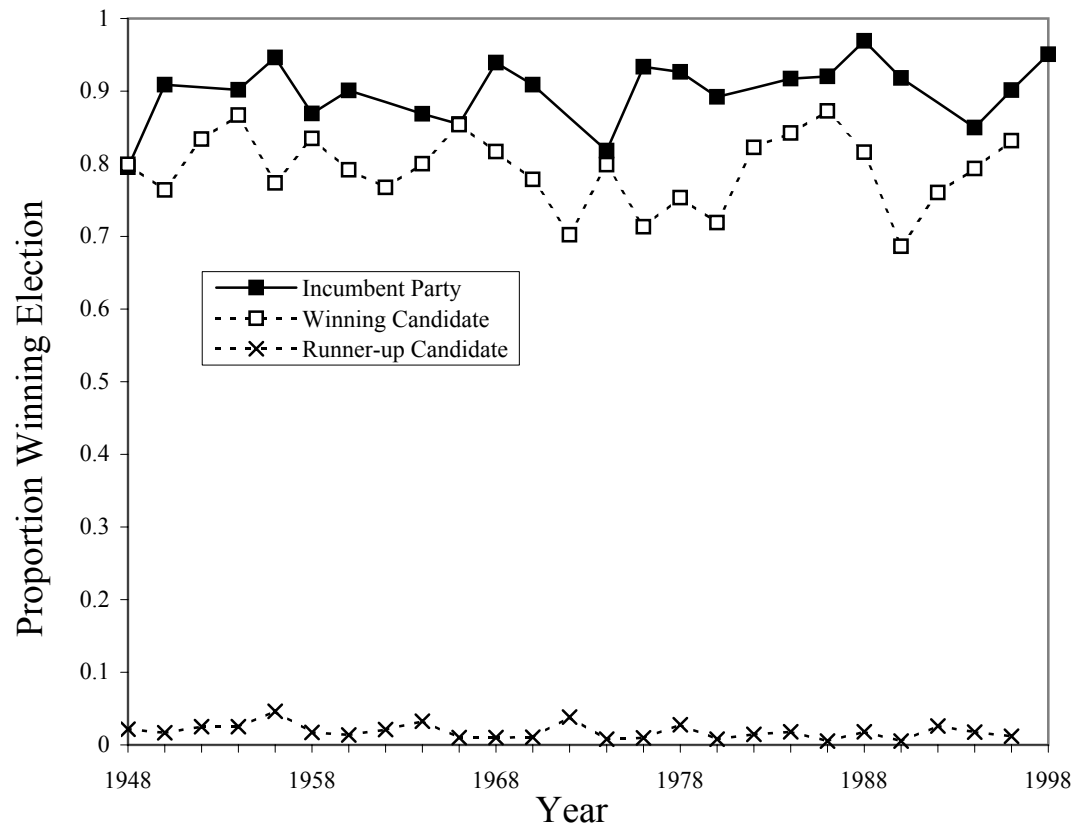


FIGURE I: Electoral Success of U.S. House Incumbents:
1948-1998



Note: Calculated from ICPSR study 7757. Details in Data Appendix. Incumbent party is the party that won the election in the preceding election in that congressional district. Due to re-districting on years that end with "2", there are no points on those years. Other series are the fraction of individual candidates in that year, who win an election in the following period, for both winners and runner-up candidates of that year.

Figure IIa: Candidate's Probability of Winning Election t+1, by Margin of Victory in Election t: local averages and parametric fit

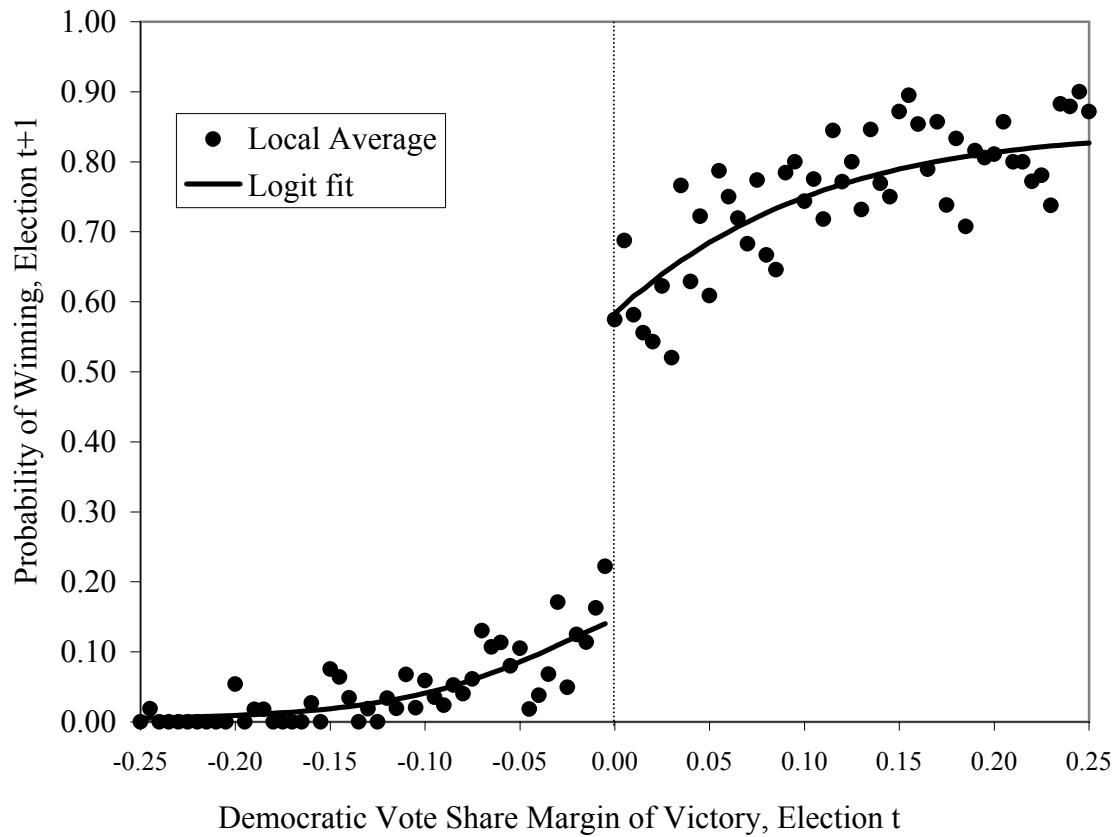


Figure IIb: Candidate's Accumulated Number of Past Election Victories, by Margin of Victory in Election t: local averages and parametric fit

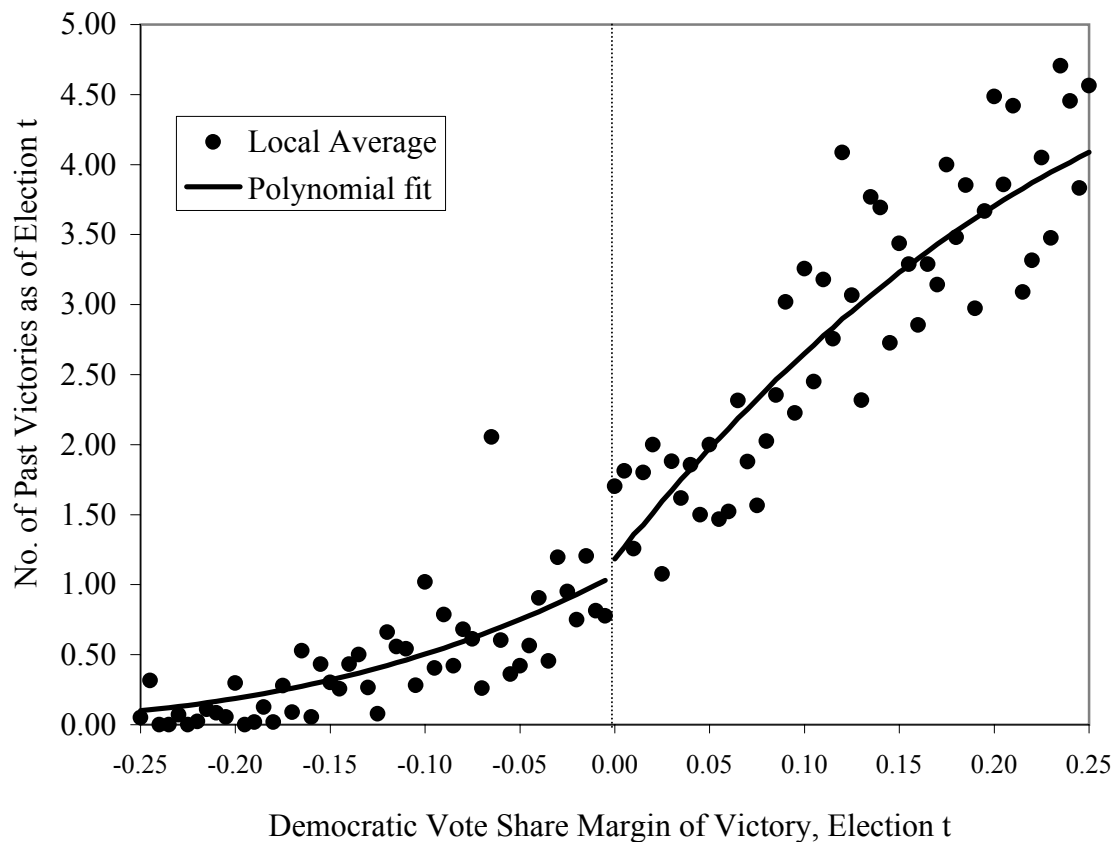


Figure IIIa: Candidate's Probability of Candidacy in Election t+1, by Margin of Victory in Election t: local averages and parametric fit

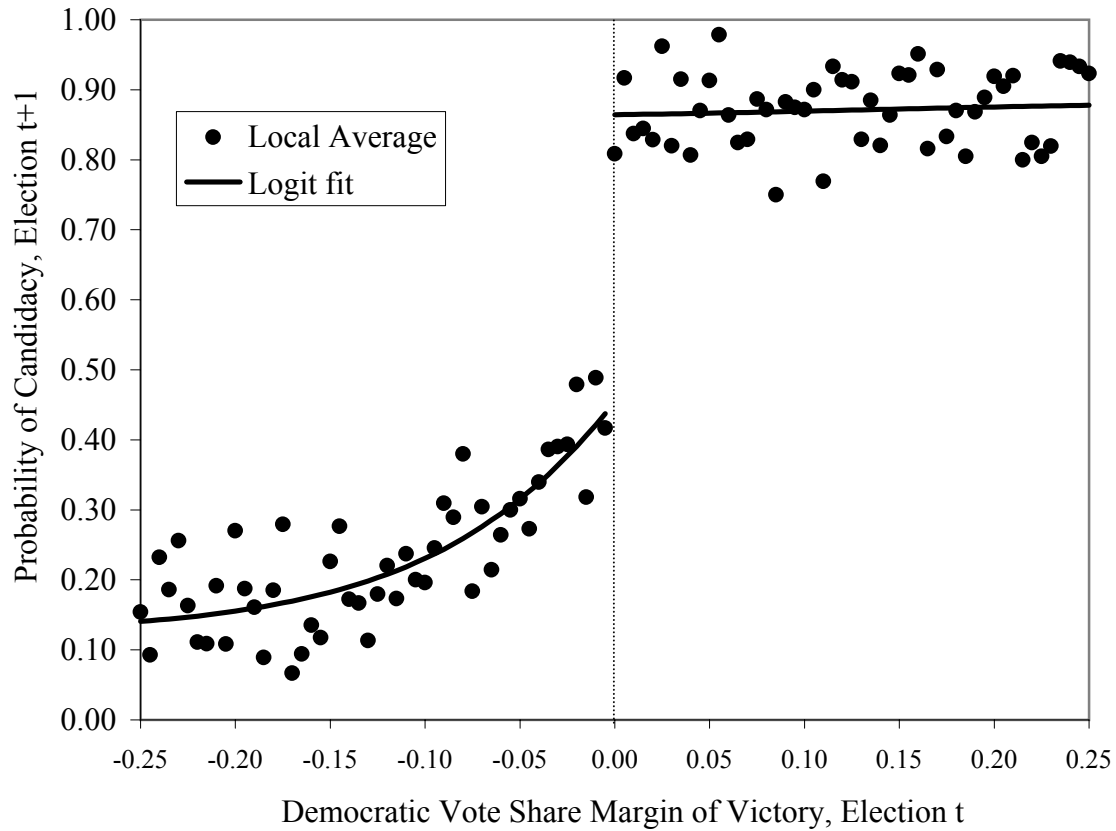


Figure IIIb: Candidate's Accumulated Number of Past Election Attempts, by Margin of Victory in Election t: local averages and parametric fit

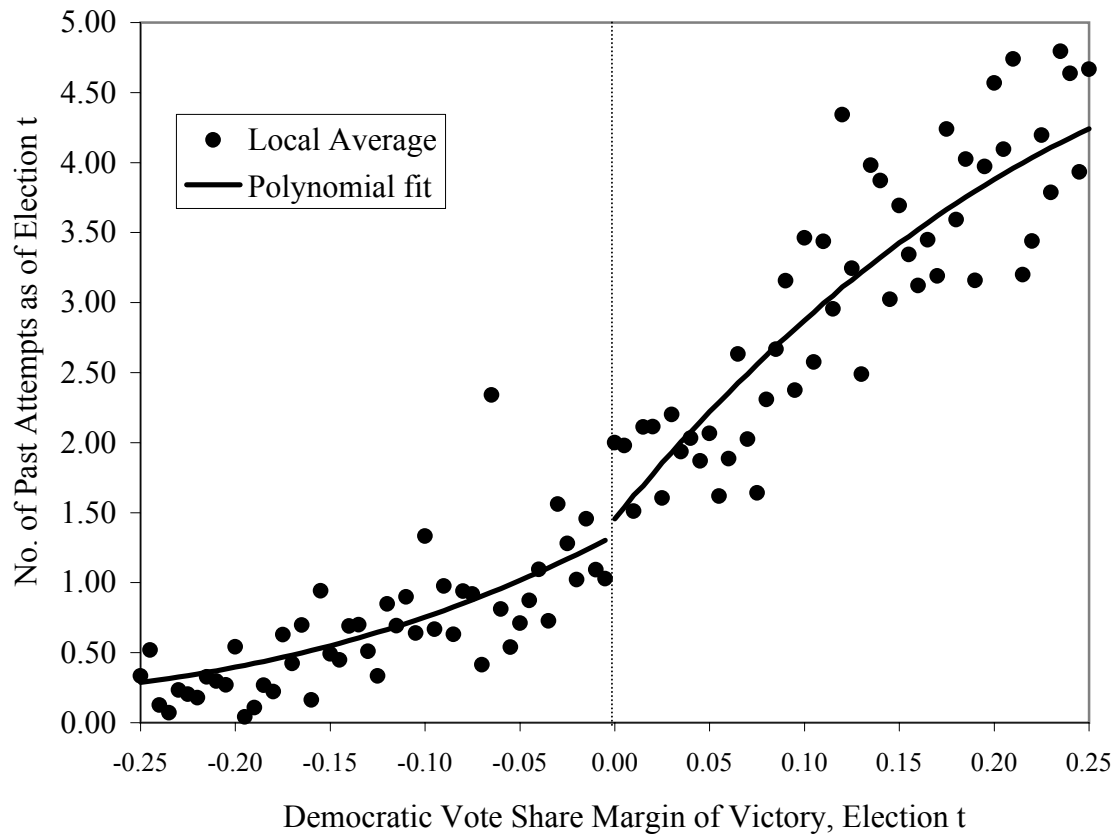


Figure IVa: Democrat Party's Vote Share in Election t+1, by Margin of Victory in Election t: local averages and parametric fit

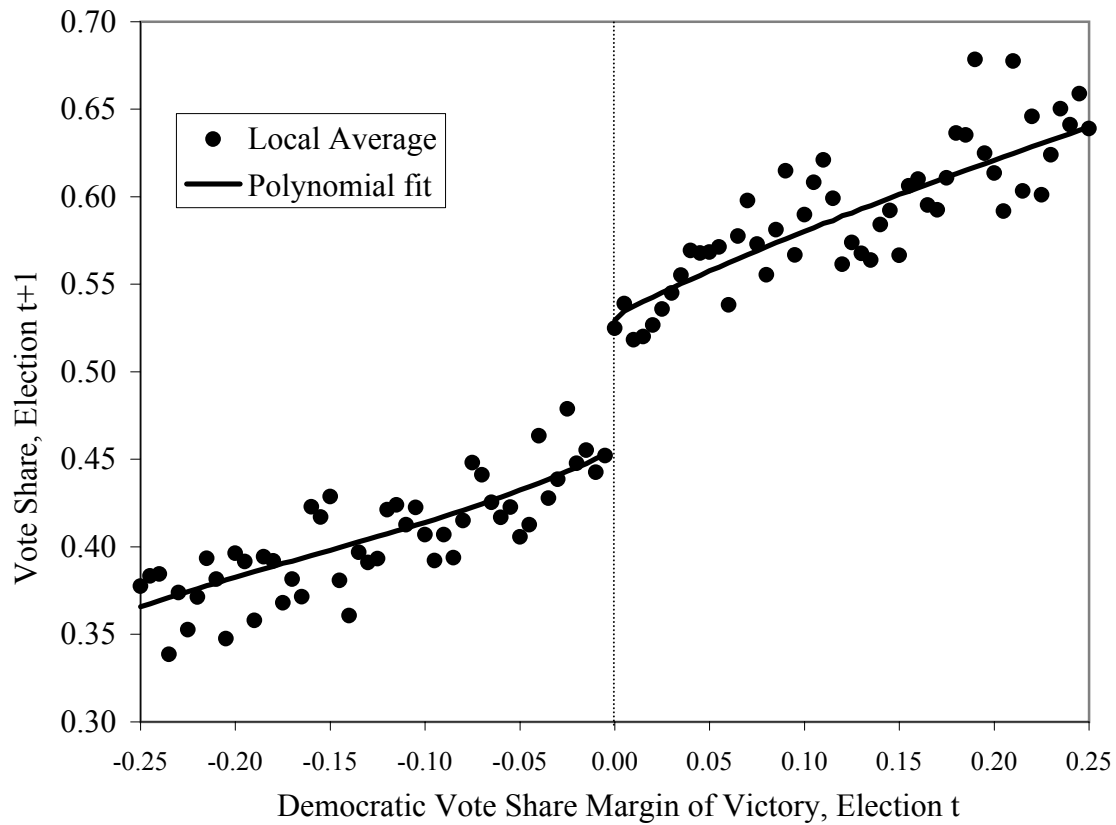


Figure IVb: Democratic Party Vote Share in Election t-1, by Margin of Victory in Election t: local averages and parametric fit

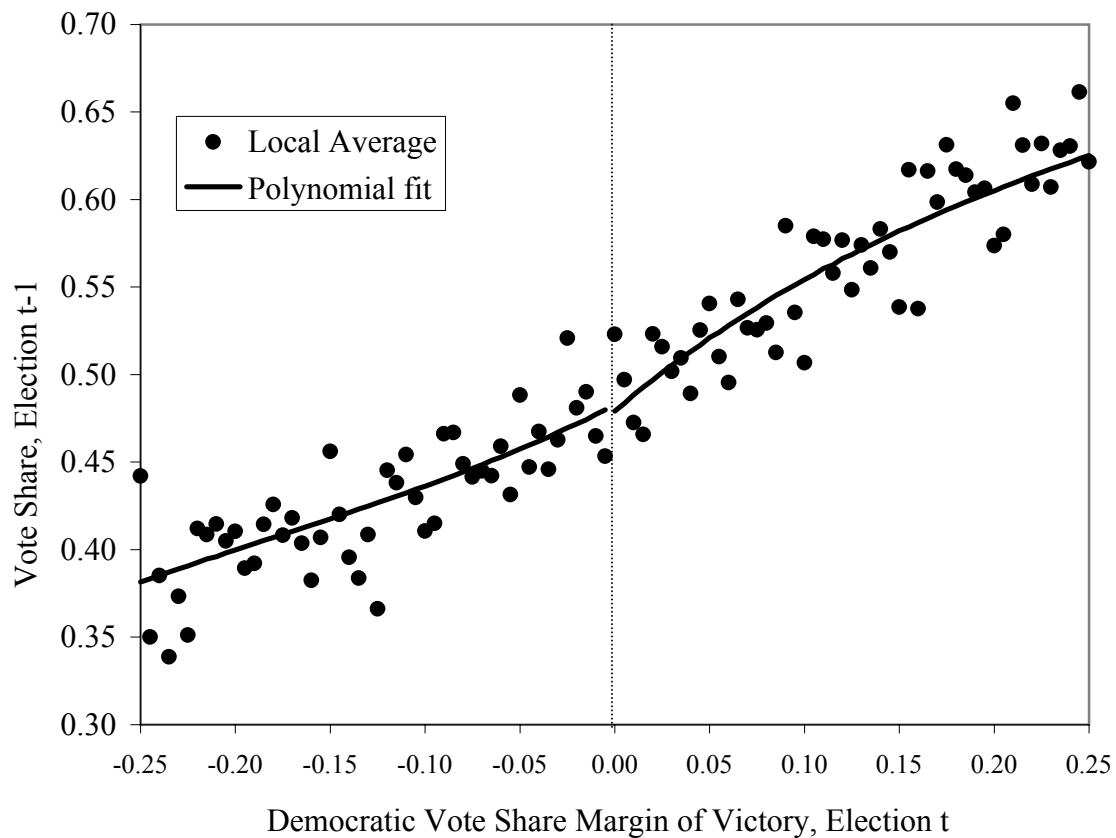


Figure Va: Democratic Party Probability Victory in Election t+1, by Margin of Victory in Election t: local averages and parametric fit

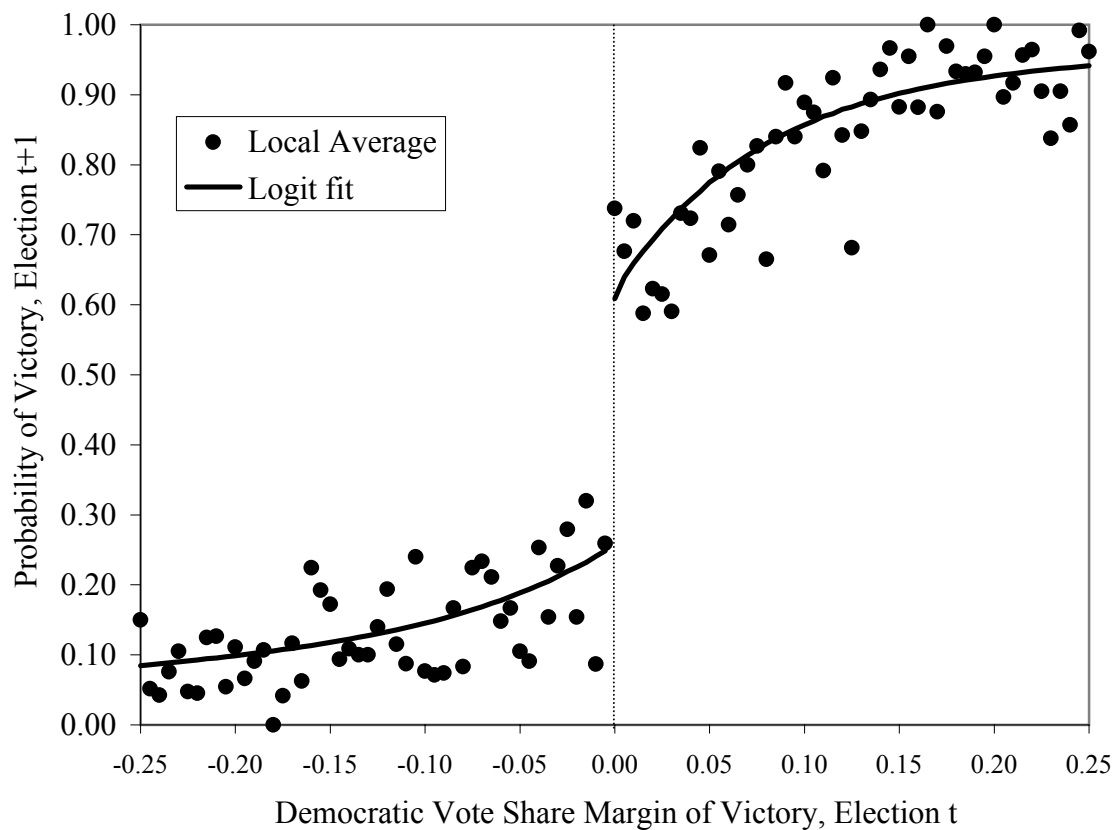


Figure Vb: Democratic Probability of Victory in Election t-1, by Margin of Victory in Election t: local averages and parametric fit

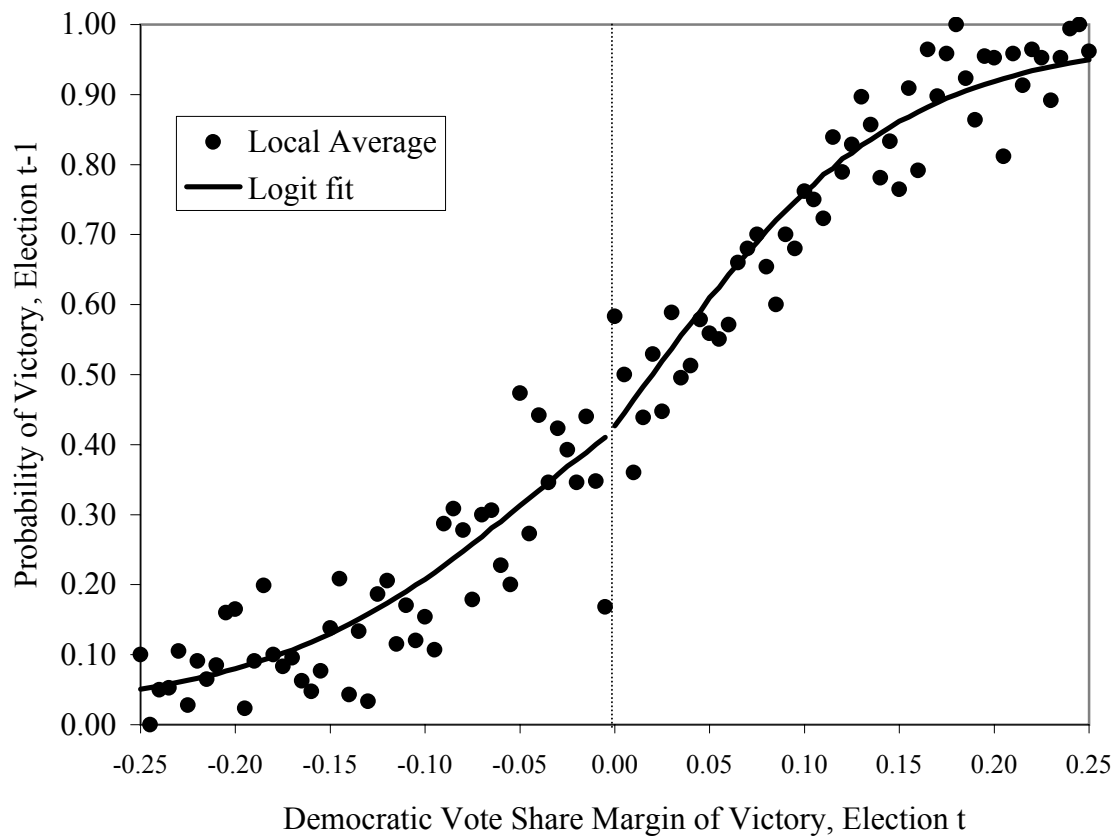


Table I: Electoral Outcomes and Pre-determined Election Characteristics: Democratic candidates, Winners vs. Losers: 1948-1996

Variable	All		Margin <.5		Margin <.05		Parametric fit	
	Winner	Loser	Winner	Loser	Winner	Loser	Winner	Loser
Democrat Vote Share Election t+1	0.698 (0.003)	0.347 (0.003)	0.629 (0.003)	0.372 (0.003)	0.542 (0.006)	0.446 (0.006)	0.531 (0.008)	0.454 (0.008)
	[0.179]	[0.15]	[0.145]	[0.124]	[0.116]	[0.107]		
Democrat Win Prob. Election t+1	0.909 (0.004)	0.094 (0.005)	0.878 (0.006)	0.100 (0.006)	0.681 (0.026)	0.202 (0.023)	0.611 (0.039)	0.253 (0.035)
	[0.276]	[0.285]	[0.315]	[0.294]	[0.458]	[0.396]		
Democrat Vote Share Election t-1	0.681 (0.003)	0.368 (0.003)	0.607 (0.003)	0.391 (0.003)	0.501 (0.007)	0.474 (0.008)	0.477 (0.009)	0.481 (0.01)
	[0.189]	[0.153]	[0.152]	[0.129]	[0.129]	[0.133]		
Democrat Win Prob. Election t-1	0.889 (0.005)	0.109 (0.006)	0.842 (0.007)	0.118 (0.007)	0.501 (0.027)	0.365 (0.028)	0.419 (0.038)	0.416 (0.039)
	[0.31]	[0.306]	[0.36]	[0.317]	[0.493]	[0.475]		
Democrat Political Experience	3.812 (0.061)	0.261 (0.025)	3.550 (0.074)	0.304 (0.029)	1.658 (0.165)	0.986 (0.124)	1.219 (0.229)	1.183 (0.145)
	[3.766]	[1.293]	[3.746]	[1.39]	[2.969]	[2.111]		
Opposition Political Experience	0.245 (0.018)	2.876 (0.054)	0.350 (0.025)	2.808 (0.057)	1.183 (0.118)	1.345 (0.115)	1.424 (0.131)	1.293 (0.17)
	[1.084]	[2.802]	[1.262]	[2.775]	[2.122]	[1.949]		
Democrat Electoral Experience	3.945 (0.061)	0.464 (0.028)	3.727 (0.075)	0.527 (0.032)	1.949 (0.166)	1.275 (0.131)	1.485 (0.23)	1.470 (0.151)
	[3.787]	[1.457]	[3.773]	[1.55]	[2.986]	[2.224]		
Opposition Electoral Experience	0.400 (0.019)	3.007 (0.054)	0.528 (0.027)	2.943 (0.058)	1.375 (0.12)	1.529 (0.119)	1.624 (0.132)	1.502 (0.174)
	[1.189]	[2.838]	[1.357]	[2.805]	[2.157]	[2.022]		
Observations	3818	2740	2546	2354	322	288	3818	2740

Note: Details of data processing in Data Appendix. Estimated standard errors in parentheses. Standard deviations of variables in brackets. Data include Democratic candidates (in election t), Democrat vote share and win probability is for the party, regardless of candidate. Political and Electoral Experience is the accumulated past election victories and election attempts for the candidate in election t, respectively. The "opposition" party is the party with the highest vote share (other than the Democrats) in election t-1. Details of parametric fit in text.

Table II: Effect of Winning an Election on Subsequent Party Electoral Success: Alternative Specifications, and Refutability Test, Regression Discontinuity Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	Vote Share t+1	Vote Share t+1	Vote Share t+1	Vote Share t+1	Vote Share t+1	Res. Vote Share, t+1	1st dif. Vote Share, t+1	Vote Share t-1
Victory, Election t	0.077 (0.011)	0.078 (0.011)	0.077 (0.011)	0.077 (0.011)	0.078 (0.011)	0.081 (0.014)	0.079 (0.013)	-0.002 (0.011)
Dem. Vote Share, t-1	----	0.293 (0.017)	----	----	0.298 (0.017)	----	----	----
Dem. Win, t-1	----	-0.017 (0.007)	----	----	-0.006 (0.007)	----	-0.175 (0.009)	0.240 (0.009)
Dem. Political Experience	----	----	-0.001 (0.001)	----	0.000 (0.003)	----	-0.002 (0.003)	0.002 (0.002)
Opp. Political Experience	----	----	0.001 (0.001)	----	0.000 (0.004)	----	-0.008 (0.004)	0.011 (0.003)
Dem. Electoral Experience	----	----	----	-0.001 (0.001)	-0.003 (0.003)	----	-0.003 (0.003)	0.000 (0.002)
Opp. Electoral Experience	----	----	----	0.001 (0.001)	0.003 (0.004)	----	0.011 (0.004)	-0.011 (0.003)

Note: Details of data processing in Data Appendix. N= 6558 in all regressions. Regressions include a 4th order polynomial in the margin of victory for the Democrats in Election t, with all terms interacted with the Victory, Election t dummy variable. Political and Electoral Experience is defined in notes to Table II. Column (6) uses as its dependent variable the residuals from a least squares regression on the Democrat Vote Share (t+1) on all the covariates. Column (7) uses as its dependent variable the Democrat Vote Share (t+1) minus the Democrat Vote Share (t-1). Column (8) uses as its dependent variable the Democrat Vote Share (t-1). Estimated standard errors (in parentheses) are consistent with state-district-decade clustered sampling.