A Regression Discontinuity Design for Studying Divided Government

Patricia A. Kirkland *
Woodrow Wilson School of Public and International Affairs
Princeton University
pak2@princeton.edu

Justin H. Phillips
Department of Political Science
Columbia University
jhp2121@columbia.edu

February 12, 2019

Abstract

The regression discontinuity design (RDD) is a valuable tool for identifying causal effects with observational data. However, applying the traditional electoral RDD to the study of divided government is challenging. Because assignment to treatment in this case is the result of elections to multiple institutions, there is no obvious single forcing variable. Here we use simulations in which we apply shocks to real-world election results in order to generate two measures of the likelihood of divided government, both of which can be used for causal analysis. The first captures the electoral distance to divided government and can easily be utilized in conjunction with the standard sharp RDD toolkit. The second is a simulated probability of divided government. This measure does not easily fit into a sharp RDD framework, so we develop a probability restricted design (PRD) which relies upon the underlying logic of an RDD. This design incorporates common regression techniques but limits the sample to to those observations for which assignment to treatment approaches "as-if random." To illustrate both of our approaches, we reevaluate the link between divided government and the size of budget deficits.

*We are indebted to Olle Folke who provided invaluable methodological advice and code. This project also benefited greatly from Liangyu Zhu, our dedicated research assistant. A prior version of this paper was presented at the State Politics and Policy Conference, June 2018. We thank Jesse Crosson, Chris Skovron, and other conference participants for helpful comments.

Perplexed by the increased frequency of divided government in the post World War II era, political scientists began studying its implications for American politics and lawmaking (Binder 2015). Led by David Mayhew's (1991) seminal book, *Divided We Govern*, much of this work has focused on legislative performance, considering whether split partisan control of the legislative and executive branches hinders the production of important bills (cf., Binder 2003; Chiou and Rothenberg 2003; Rogers 2005; Hicks 2015). Others, though, have considered the effects of divided government on a variety of additional outcomes, including the size of budget deficits (McCubbins 1991), presidential approval (Nicholson, Segura, and Woods 2002), the ability of executives to achieve their policy aims (Kousser and Phillips 2012; Howell 2003), the timeliness of legislative action (Moraski and Shipan 1999; Binder and Maltzman 2002), and the frequency of investigations of the executive (Kriner and Schwartz 2008). Given its ubiquity, unpacking the consequences of divided government is fundamental for understanding separation of powers systems.

Despite decades of research, isolating the effects of split partisan control remains a challenge. Obviously, divided government is not randomly distributed, and places that routinely experience split party control may be different from those that tend not to. Indeed, Figure 1 illustrates the frequency with which individual states experience divided government varies quite dramatically and there are, in some cases, stark regional differences.[1] If the characteristics that cause a place to "select into" divided government are systematically related to outcomes, we must worry about the potential for biased results.

Studies of divided government have typically tried to address such concerns by including covariates in regression models. These allow researchers to account for between-unit differences that are both observable and measurable. To account for unobserved, but time-invariant, differences researchers sometimes include fixed effects. In practice, however, it is unclear which covariates ought to be included in statistical models, in large part because researchers are interested in ex-

[1]The states that have experienced divided government most frequently between 1968 and 2010 are Alaska (84% of the time), Delaware (81%), Michigan (81%), and New York (81%). The states that have experienced divided government least frequently are Georgia (5%), Maryland (12%), Hawaii (19%), and South Dakota (19%). It is also worth noting, that the south is the region of the country that has been the least likely to experience split partisan control of government (though there are exceptions, such as Tennessee (65%) and Virginia (63%).

plaining the outcomes of complex processes. Moreover, the determinants of outcomes may change over time or across policy areas, limiting the value of fixed effects. While we have undoubtedly learned a great deal from research on divided government that has used these empirical strategies, studies have at times yielded conflicting results (cf., empirical studies of legislative productivity). Such inconsistencies may stem from differences in how researchers address challenges of model specification as well as the statistical techniques and research designs that they are able to employ.

Increasingly, political scientists have turned to regression discontinuity designs (RDDs) to estimate causal effects using observational data. An RDD allows researchers to compare outcomes across units that are quite similar in terms of their probability of being treated but differ only in whether the treatment is applied. This mitigates the threat of endogeneity, particularly concerns of selection bias and omitted variable bias (Lee and Lemieux 2010). Furthermore, compared to other non-experimental approaches, the identification assumptions of an RDD are transparent and more readily testable (Lee and Lemieux 2010). To date, RDDs have been utilized to estimate the causal effect of a variety of non-randomly assigned political treatments, such as incumbency, partisanship, and ideology (e.g., Lee, Moretti, and Butler 2004; Lee 2008; Gerber and Hopkins 2011; Broockman 2014; Hall 2015).

In these and many other applications, treatment status is the result of a single type of election. Gerber and Hopkins, for example, study whether the partisanship of a mayor affects municipal fiscal policy, while Broockman (2014) considers how electing female state legislators influences women's subsequent political participation. In such settings, candidate vote share is readily operationalized as the forcing variable—i.e., the variable that determines assignment to treatment.

However, applying this design to the study of divided government is not straightforward. Because divided government is the result of multiple election outcomes—elections to the governorship, the state assembly, and the state senate—there is no obvious forcing variable. Legislative seat shares, for example, cannot easily be combined with gubernatorial election results since they each measure distinct quantities of interest. Furthermore, the ubiquity of gerrymandered legislative districts and uncontested races means that closely divided chambers (in terms of the partisan distribution of seat shares) are not necessarily chambers where majority status is up for grabs. We suspect that challenges in constructing a forcing variable have inhibited the use of RDD in studies of divided government.

Here, using historical data on state-level elections, we offer and implement a simulations-based approach for addressing these challenges. In order to do so, we build upon our prior work (Kirkland and Phillips 2018) as well as work by Folke (2014) and Fiva, Folke, and Sørensen (2018). Our approach is to apply shocks to three types of elections—elections to the state senate, state assembly, and the governorship—to create two measures that capture the underlying (but ultimately unobservable) likelihood of divided government.

The first measure, which we refer to as electoral *distance to divided government*, is the size of the smallest state-level vote shock that would have produced a different outcome (in terms of unified or divided government) than the outcome we actually observe. This measure is analogous to the vote share measures commonly employed in existing electoral RDDs and can readily be utilized in the standard sharp RDD analysis.[2]

The second measure, the *simulated probability of divided government*, is operationalized as the proportion of simulations for a given state year in which neither the Democratic nor Republican party wins full control of government. This measure may be more intuitively appealing than the traditional vote share forcing variable. Arguably, a probability measure (if properly constructed) should be better at identifying those observations where the likelihood of receiving treatment is truly close to 50-50. This would be the case, for example, if two observations share an identical value of an electoral forcing variable, say 5%, but differ in the likelihood that a 5% vote shift would occur. However, the application of any probability measure to the traditional RDD set

[2]Our approach is also similar to emerging work that uses a "centering" procedure to collapse multiple forcing variables into a single binding score. This procedure selects (from the set of forcing variables) the smallest value that would have changed treatment assignment. Such an approach is utilized by Folke (2014) in his study of the influence of parties in Sweden's proportional representation system. In this application, the collapsed forcing variable is the smallest shift in a party's vote share that would change its allocation of seats in municipal councils. Our simulations generate measures of the distance to divided government that are quite similar to those we would calculate using a centering procedure, but we think there are two key advantages of our approach. First, our simulations also produce a measure of the probability of divided government which we use in a probability restricted design. Second, the centering option would require the assumption that shifts in legislative vote-shares would be uniform across districts while the simulations allow electoral shocks to vary across districts.

up is complicated somewhat by the fact that it is not deterministic. That is, not all states with a probability of divided government over 50% will actually experience divided government (and vice versa). This is most likely to happen for state years in which the probability of divided government is at or near 50% (i.e., those observations that would be of primary interest in an RDD).[3]

To make use of this measure, we depart from a typical RDD application, but still rely upon its underlying intuition. Specifically, we estimate OLS regressions that use only observations that have close to a 50% probability of receiving the treatment. By restricting the sample in this way, we have a set of observations that have nearly the same probability of experiencing divided government, i.e., assignment to treatment is “as-if random.” Doing so substantially mitigates the threat of endogeneity and allows us to estimate the effect of divided government as we would if we were able to randomly assign it. Although we modify the traditional RDD, this strategy is somewhat similar to one employed by Anzia and Berry (2011) and is certainly in the spirit of the traditional approach, and like a textbook RDD, it allows us to avoid concerns of selection bias. For ease of explication, we refer to this RDD-inspired approach as a “probability restricted design” or PRD.

To demonstrate these two simulations-based approaches to studying divided government, we proceed with an application. Specifically, we reevaluate the disputed claim that divided government increases (decreases) the size of budget deficits (surpluses). State budgeting is an ideal arena for applying our approach (and applying RDD in general) because: (1) state budgets are substantively important; (2) data on state fiscal policy and fiscal outcomes have been systematically collected for decades, meaning that there are sufficient data for an RDD; and (3) budgeting and its associated fiscal outcomes are determined by complex processes where establishing causal relationships with more traditional observational methods is difficult. For our application we use historical data on state budget surpluses/deficits from 1968-2010. In doing so, we find evidence of a causal relationship—divided government leads to higher deficits (lower surpluses). We extend our analysis to consider whether this result is due to changes in revenues, expenditures, or both.

[3]Although our probability measure is conceptually different from common forcing variables, we refer to it as a forcing variable because it is a proxy for the underlying probability of assignment to treatment.

Our results indicate that the "culprit" is an increase in spending that is not mirrored by increases in revenues.

Overall, our study shows that RDD (or RDD-inspired approaches) are indeed promising tools for the study of divided government, especially at the level of the U.S. states. The toolkit that we develop and demonstrate here can readily be used by others to further examine the consequences of divided government, or it can be adapted to explore additional central questions such as the impact of partisan control of government. Our toolkit should open new doors for using the techniques of causal inference to study separation of powers systems.

1 The Challenge of Regression Discontinuity Designs & Divided Government

Many fundamental questions in political science involve treatments—such as democracy, partisanship, race, or gender—that cannot be randomly assigned, making the use of experimental techniques implausible. Divided government is one of these. Using observational data, however, researchers often can draw on other techniques to identify causal effects (see Angrist and Pischke 2008 for a review). One such technique is a regression discontinuity design (RDD). Essentially, an RDD allows researchers to compare outcomes across units that are quite similar in terms of their likelihood of being treated but differ only in whether the treatment is actually applied. To execute an RDD, researchers must find or construct an appropriate forcing variable—the variable that determines assignment to treatment.

Because the design itself addresses selection bias, researchers do not need to account for potential confounders (covariates) to estimate causal effects, although covariates may be incorporated to increase the precision of estimates. In contrast, traditional multivariate regression produces unbiased results only if models include all potential confounding variables, a crucial assumption that is difficult or even impossible to test. A challenge arises because it is often unclear what potential confounders ought to be included, especially given the complex processes that political scientists study. This is evident in the empirical divided government literature where the set of covariates employed in regression models tends to vary widely across studies, even those that utilize the same or a similar dependent variable. Differences in model specification may lead to conflicting findings, as we observe in the literature that studies the effect of divided government on legislative productivity (cf. Mayhew 1991; Gray and Lowery 1995; Krehbiel 1998; Squire 1998; Coleman 1999; Bowling and Ferguson 2001; Binder 2003; Rogers 2005).

Although researchers diligently try to account for relevant variables, the possibility of unknown or unmeasurable determinants also raises the spectre of omitted variable bias. One option for addressing unobserved or unmeasured differences is the use of fixed effects. However, fixed effects cannot account for unobserved variables that change over time or the possibility that the implications of fixed state characteristics themselves change over time.

A key advantage of the RDD, especially when compared to other strategies for causal inference with observational data, is that its identification assumptions are transparent and readily testable (Lee and Lemieux 2010). The primary assumption is that potential outcomes are smooth across the discontinuity in the forcing variable, that is units cannot select or sort into treatment or control. As long as this assumption is valid, one can be reasonably sure that differences in characteristics are evenly distributed among treatment and control units, at least among those observations located close to the threshold or cutpoint between treatment and control.

While an RDD supports causal inference, a limitation of the design is that its results apply only to a subset of observations. That is, an RDD estimates a narrow quantity of interest. Unlike traditional multivariate and fixed effects models, it produces a local average treatment effect—i.e., the effect of treatment on units that lie close to the threshold. In our case, for example, it estimates the effect of a state moving from barely having unified to barely having divided government. In contrast, traditional multivariate models estimate an average treatment effect, while fixed effects approaches estimate within-unit effects. For researchers specifically concerned with causal inference, this trade off is sensible.

RDD has found widespread use in political science to study representation. These efforts have often explored the characteristics of lawmakers that may affect their policy choices while in office or some other political outcome of interest, such as the size of the incumbency advantage or rates of participation in future elections. The characteristics studied include partisan affiliation, race, gender, occupations, and ideology (Gerber and Hopkins 2011; Broockman 2014; Ferreira and Gyourko 2014; Hopkins and McCabe 2012; Erikson and Titiunik 2015; Hall 2015; Kirkland n.d.). In these applications, a single type of election determines treatment status, and there is a clear threshold—the candidate with the most votes wins. This makes the creation of a forcing variable quite easy. For example, Ferreira and Gyourko study the effect on fiscal policy of electing a woman mayor by focusing on elections in which a female candidate faces a male opponent. The forcing variable in their study is simply operationalized as the female candidate's vote share, and when a woman wins more than 50% of the votes, the city is exposed to treatment.

Constructing a forcing variable is trickier, however, when political treatments are the result of elections to multiple institutions as is common in separation of powers systems. Divided government in the U.S. states, the focus of our application, is a prime example of such a treatment. Because this treatment is jointly determined by elections to the state assembly, the state senate, and the governorship, there is no direct measure to serve as a forcing variable. Indeed, legislative seat shares and gubernatorial votes cannot be easily combined because they measure different quantities of interest.

Similarly, seat shares themselves can be a misleading indicator, given the prevalence of partisan gerrymanders and uncontested elections. A closely divided legislative chamber, in terms of the partisan distribution of seat shares, does not necessarily mean that the chamber was or is at high risk for a different outcome (e.g., control by the minority party). For example, in 1999 both Texas and Tennessee had closely divided senates creating the appearance that the majority party's status was uncertain. However, neither state had any highly competitive races, and no state senator had a winning margin of less than 10 percentage points. Moreover, nearly half of the seats were uncontested. As a result, we need to know much more about these elections than just the number of seats each party ultimately won.

2 Electoral Simulations

To create the measures that support our causal analyses, we build upon the simulations developed in our prior work (Kirkland and Phillips 2018). Specifically, we apply shocks of varying magnitudes to real-world electoral results in order create measures that capture the underlying likelihood of divided government. Unlike our prior efforts, where shocks were drawn from a uniform distribution (ranging from -1 to 1), here we utilize a separate distribution for each state that is informed by that state's historical election results. Doing so ensures that the shocks used in our simulations are similar in magnitude to the shocks likely to occur in each state.[4]

Indeed, data on election results justify our decision not to simply draw shocks from a uniform distribution (-1 to 1). Figure 2 plots for 49 states (Nebraska is excluded due to its nonpartisan legislature) the distribution of historical shocks to Democratic vote share in elections to the state assembly, senate, and governorship. The shock to Democratic vote share in any given office is simply the difference between the Democratic candidate's vote share in the current and prior election. For example, the Democratic candidate for governor of Connecticut in 1974 received 58.3% of the vote, up from 46.2% of the vote in 1970. This means that the historical Democratic vote shock for the 1974 gubernatorial election is 12.1 percentage points. To be clear, Figure 2 does not plot a state-level measure or a state's office-level average, but rather treats shocks to all a state's assembly, senate, and gubernatorial elections as separate observations. Each state's panel plots the distribution of the historical shocks from all of the elections in our dataset. For example, the figure for California plots 1,983 separate shocks.

Nearly all states exhibit something close to a normal distribution centered around zero. The main difference that we observe across states is that some have a flatter distribution of shocks, i.e., their distributions have greater variance. And, these are the states that tend to experience larger electoral shocks. For example, compare the distributions for California, which has a mean

[4]By relying on historical election results, we expect our simulations to produce a more accurate measure of our "probability of divided government" forcing variable (assuming that the size of the typical electoral shock varies across states). However, this decision is unlikely to have much of an effect on the values we generate for the "distance to divided government" forcing variable, since that measure simply captures the smallest electoral shock that would produce a different outcome and not the likelihood of that shock actually occurring.

of 0.006 and a standard deviation of 0.113, to Massachusetts, which has the same mean but a standard deviation of 0.150. Since 1968, only 25% of the legislative and gubernatorial shocks in California have been greater than 10 percentage points, while in Massachusetts over 43% of the shocks fall into this category.

The key data for generating our forcing variables are historical state election results. For state assembly and state senate elections, we rely on ICPSR's "State Legislative Election Returns (1967-2010)" dataset (Klarner et al. 2013). These data include candidates' names, party affiliations, and vote counts by state legislative district. We supplement these with gubernatorial election returns from *Congressional Quarterly's* (2003) "Voting and Elections Collection." We utilize gubernatorial and legislative election data to calculate historical electoral shocks. These data, along with Dubin (2007), are also used to determine the partisan distribution of legislative seat shares.

The first step in each simulation is to establish the size of a state-level electoral shock (S_i), the value of which then constrains the size of the district-level shocks (ΔV) that we ultimately apply to real world election results. S_i is randomly drawn from a normal distribution with a mean and standard deviation that are equal to the mean and standard deviation of the actual distribution of historical aggregate election results for state i. The value of S_i can be either positive or negative, and smaller (larger) values of S_i produce smaller (larger) values of ΔV.

The second step is to take, for each legislative district (j) in the state (i), a new random draw (D) from a normal distribution with the mean and standard deviation of historical election shocks from each type of district in the state. For example, electoral shocks for a state senate seat in California, D_{ij} will be drawn from a normal distribution with the mean of historical shocks across all state senate districts in California. By incorporating (D), we allow for random variation in the size of shocks across districts.[5] Each ΔV_{ij}, then, is a straightforward function of these two draws:

[5] If, for each district, we were instead to draw D_{ij} from a normal distribution with the mean and standard deviation of that district's historical election shocks during a given redistricting period (using redistricting data from Klarner 2018), our forcing variable would be virtually identical. We do not take this approach only because there are often very few data points from which to generate a district's distribution.

$$\Delta V_{ij} = S_i + S_i * D_{ij} \tag{1}$$

To calculate the vote shocks that we apply to gubernatorial elections, we simply take a random draw from a normal distribution with the mean and standard deviation of the distribution of historical gubernatorial election shocks for that state.

The third step of each simulation is to apply the shocks. In every legislative district election and gubernatorial election, we add ΔV_{ij} to the Democratic candidate's vote share while subtracting ΔV_{ij} from the Republican's vote share. We then determine which candidate wins. We translate our simulated election results into legislative seat shares and combine these with the simulated gubernatorial election results to determine the partisan control of state government.

After each simulation, we record not only whether the the simulated election results produced divided or unified government but also whether the simulated outcome differs from the actual observed outcome. We repeat the simulation process 40,000 times for all state-election years, using the results to identify the smallest state-level vote shock (S) that produces a different outcome in the majority of simulations. This measure, which we refer to as the electoral distance to divided government, then becomes our first forcing variable.

To generate our second forcing variable—simulated probability of divided government—we simply calculate the proportion of simulations for a given state year in which neither political party wins unified control of government. For example, if a given observation has a proportion of 0.5, it means that the state was as likely to experience divided as unified government. Accordingly, a value of .95 implies that the state was almost certain to experience divided government, while a much lower proportion of .05 indicates state government was almost certain to be controlled by one party. We discuss the distribution of this variable in much greater detail in Section 2.2.

We run these simulations for all states from 1968 though 2010, with the exception of five—Arizona, Louisiana, Nebraska, New Jersey, and North Dakota. Nebraska is excluded due to its use of nonpartisan legislative elections. Louisiana is omitted because its jungle primary system leaves few legislative elections that are competitive two-party contests. Finally, the use of multimember districts forces us to exclude Arizona, Nebraska, and New Jersey. These states have multimember district elections with two common features that are incompatible with our simulations. First, multiple legislators are elected simultaneously in the same district. Second, one entire chamber of the legislature is elected in this type of multimember district.

2.1 Simulated Distance to Divided Government

The distribution of our distance to divided government forcing variable is displayed on the x-axis of Figure 3 where each bin represents 0.025 percentage points. Observations of the forcing variable to the left of the cutpoint, i.e., those with negative values, are assigned to unified government while observations to the right of the cutpoint (positive values) are assigned to divided government. Observations that lie close to the cutpoint are states which were about as likely to experience unified or divided government. Substantively, a forcing variable of -0.04 indicates that the state in question experienced unified government and that a state-level vote shift of 4 percentage points or greater (to the opposition party) would have produced divided government. Correspondingly, a value of 0.10 means the state experienced divided government but that a state-level vote shift of 10 percentage points or greater would have produced unified partisan control.

Note that a large number of observations (245 in total) fall at the extreme ends of the distribution. These are observations for which our simulations do not produce a value of the forcing variable. In these state-years a very large shock would be required to produce a different electoral outcome, and, draws from a normal distribution, even one informed by the state's electoral history, do not uncover a shock large enough to produce such an outcome, at least not in a majority of simulations. Keep in mind, large shocks/vote swings are relatively uncommon.[6] We artificially place observations with no value of the forcing variable at either the far left or far right of the

[6]One such observation is New York 1996. Elections that year produced divided government, with a Republican (George Pataki) as governor, Republican control of the state senate (with a 35-26 majority), and Democratic control of the state assembly (with a 96-54 majority). Because Gov. Pataki was not up for reelection, the only way the state could have experienced a different outcome (in terms of unified or divided control of government) was if Republicans gained a majority in the assembly. However, this would have required a whopping electoral gain of 42 seats. In order to win the 42nd most competitive Democratic assembly seat, the Republican candidate would have required a vote shock of at least 48 percentage points. A shock of this size is highly improbable. Indeed, when we generate our estimated of probability of divided government for each observation in our dataset (which we discuss in greater detail in the following section) the value for New York 1996 is 100%.

distribution. Observations placed at the left end are those with unified government and those at the right end are those with divided government. That these state-years lack a value of the forcing variable is inconsequential for any subsequent RDD analysis. This is because such observations would fall outside reasonably-sized bandwidths and therefore would not be utilized in the local linear regression models that are commonly used to test for the existence of a causal relationship.

Figure 4 presents additional descriptive statistics, this time focusing on cross-state variation. The first panel shows the average magnitude of the forcing variable by state. Lower values here indicate states where the partisan control of government was more frequently uncertain. There is a fair amount of variation across states. The overall mean is 10%, with values ranging from a low of 4% (Michigan) to a high of 16.5% (Arkansas).

The second panel reports the share of state observations for which the forcing variable falls within a 5% window around the cutpoint. We display a 5% window here for two reasons. First, there is a general consensus in studies that utilize electoral RDDs that elections decided by 5 points or fewer are "close" elections. Second, the optimal bandwidth in our main analysis (which we will discuss in greater detail below) is also just a shade over 5%. Observations that fall within this window form the core of our RDD analysis. Importantly, there is no state that scores a zero on this measure which means that all states have at least some election cycles where we could think of the partisan control of government as narrowly decided. That being said, there is also fair amount of cross-state variation. While the average value across states is 28%, there are six states (Alabama, Iowa, Michigan, Montana, Pennsylvania, and Washington) where 50% or more of the observations fall within 5% of the cutpoint. On the other end of the spectrum, there are six states (Arkansas, Connecticut, Kentucky, South Dakota, Virginia, and Wyoming) where fewer than 15% of observations fall within this window.

Overall, we find these descriptive statistics reassuring. First, within commonly-employed RDD bandwidths (i.e., 5 or 10%), there are a large number of observations (189 and 376 respectively), meaning that we should be able to generate relatively precise estimates. Second, within these bandwidths all states will contribute to our analysis. This means that although we "only" estimate a local-average treatment effect, we are not worried that this effect reflects outcomes in just a handful of states.

This forcing variable is analogous to vote share measures commonly employed in electoral RDDs. Because it has a sharp discontinuity (in this case at zero), it can easily be utilized in conjunction with the standard sharp RDD toolkit.

2.2 Simulated Probability of Divided Government

The distribution of our second forcing variable is displayed in Figure 5. Here the x-axis is the estimated probability of divided government, while the y-axis is a count of the number of state years that fall into each bin (again bins represents 0.025 percentage points). Those observations that lie near 50%—the midpoint of the distribution—are those that are nearly as likely to experience divided as unified government. Observations that fall at the far right-hand side of the distribution are almost certain to experience divided government, while those at the opposite end are almost certain to experience unified government. Approximately 60% of our observations fall into one of these two categories.

Figure 6 presents further descriptive statistics. The first panel shows the average probability of divided government for each state across our entire time period. Importantly, there is considerable variation, which speaks to the concern about selection bias in observational studies of divided government—not all states are equally susceptible to split partisan control. The mean probability across all states is 53%, ranging from a low of 13% for Georgia to a high of 84.6% for New York.

The second panel displays the share of state observations for which the forcing variable falls within ± 5 percentage points of the midpoint. Like the second panel of Figure 4, presenting these data provides insights as to which states will contribute observations to our subsequent analyses. Here we utilize 5% to remain consistent with the descriptive statistics shown for our distance to divided government measure. Over half of all states have one or more election year within this window (4% of all observations fall within this window, this compares to 28% for the distance to divided government forcing variable). If we expand the bandwidth to 10%, then 7.2% of our observations fall within the range, with all but 12 states contributing observations.

Clearly, this measure is more conservative than our distance to divided government forcing variable, in the sense that it places fewer observations where assignment to treatment is uncertain. Of our two forcing variables, the probability measure is the most sensitive to the particular distribution from which electoral shocks are randomly drawn. As described above, for each state we draw from a normal distribution informed by historical electoral shocks in that state. The majority of such shocks tend to be small—-the mean of the distribution of historical shocks in our data is -0.07 percentage points. Correspondingly, random draws from such distributions will tend to produce small shocks. If, on the other hand, we were to make use of a uniform distribution, we would more frequently draw larger (positive or negative) shocks. This would induce more electoral volatility and place more observations closer to the midpoint and fewer in the tails. While others may disagree, we prefer the normal distribution because it closely resembles the distribution of real-world electoral shocks (recall Figure 2 above). An alternative that results in more observations near 50 percent probability would imply a level of electoral volatility that is rare in American state politics, at least during the period we study.

Additionally, it is important to keep in mind that while both of our measures aim to capture the true and unknowable probability of divided government, they are each somewhat conceptually different. The distance to divided government simply identifies the smallest shock that will produce a different outcome in terms of unified or divided government regardless of how common a shock of this magnitude might be. This is, in essence, what existing electoral RDDs do. Our simulated probability measure, on the other hand, tells us how likely it is that a particular election will produce divided government given the pre-election partisan composition of government and the size of electoral shocks that are typical in that state.

Despite these conceptual differences, when we compare these two measures, we observe a strong correlation of 0.94. Figure 7 plots distance to divided government on the x-axis and our estimated probability of divided government on the y-axis. As expected, observations for which the distance to divided government forcing variable is close to the cutpoint (0) also have values of the probability of divided government close to 50%. This high correlation increases our confidence that both forcing variables approximate the true underlying probability of divided government.

The inherent challenge in utilizing our probability measure is that in contrast to our distance forcing variable, it has no sharp discontinuity in assignment to treatment. In other words, there is no clear threshold at which we would would expect ex ante to see a discontinuous jump in the likelihood of experiencing divided government. Logically the most natural threshold would be at a 50% probability. However, among observations just below this potential threshold we should observe about half with divided government and half with unified government. The same should also be true for observations that lie just above a 50% probability.

It might be tempting to view this as a compliance problem and to address it by implementing what is often referred to as a "fuzzy" RDD (cf., Hahn, Todd and van der Klaauw 2001).[7] A fuzzy RDD uses treatment assignment (i.e., whether or not a unit is expected to be treated based on its value of the forcing variable) as an instrument for treatment. However, there are two concerns with this approach. The first is that it estimates an even narrower quantity of interest—the effect of treatment on the subset of observations that comply with treatment assignment. Unsurprisingly, this can also lead to fairly noisy estimates. While this concern is not fatal, the second is. Fuzzy RDDs (like their sharp counterparts) *still* rely upon a forcing variable where the likelihood of receiving treatment is discontinuous at a known threshold. The difference, however, is that in the fuzzy context there are additional variables (typically unobserved by the researcher) that determine whether an observation above (or below) that threshold actually complies/receives the treatment.[8]

We take advantage of the ambiguity in treatment assignment that occurs around the simulated probability of 50% to develop a probability restricted design (PRD) for estimating the causal

[7]Here, noncompliance would be defined as a state experiencing divided (unified) government even though it has a probability of divided government under (over) 50%.

[8]A classic example of a successfully implemented fuzzy RDD is a study by Van der Klaauw (2002). They employ this technique to estimate the effect of college financial aid awards on student enrollment decisions. A formula assigns points to admitted students (creating an "ability index") based on test scores and high school grade point average. Students with an ability index above certain point thresholds are offered larger financial aid awards (that is, if a student's ability index exceeds a given threshold, she is assigned to the treatment of a larger aid package). However, admissions officers can ultimately adjust the size of the final aid package based on additional student characteristics (though they cannot add more points to student's initial score on the index). This means that while researchers know in advance that the size of financial awards will be discontinuous at given thresholds, the final award package is not a deterministic function of a student's ability index, making a sharp RDD unsuitable.

effect of divided government. To do so, we estimate standard OLS regression models with divided government as the explanatory variable, but in the spirit of an RDD, we limit our sample to those observations for which the odds of experiencing divided government are very close to 50-50. By restricting the sample in this way, we create a subset of observations that have nearly the same probability of divided government but differ in whether they actually experience split partisan control—i.e., assignment to treatment approaches "as-if random." We analyze outcomes for the restricted sample in much the same way we would approach the results of an experiment where treatment was randomly assigned with a known probability.

It is important to note that the coefficient of interest in each approach is a slightly different quantity. Because standard RDD estimation weights observations based on their proximity to the threshold (with observations beyond the bandwidth weighted as zero), the coefficient on the key independent variable captures the intercept shift at the cutpoint. In contrast, because the PRD does not utilize weights, the coefficient on the key independent variable captures the average effect within the bandwidth. This difference suggests that PRD most plausibly uncovers a causal effect when we focus on observations where the odds of assignment to treatment are very close to 50-50.

As far as we know, we are not the first to adopt this sort of modification to a traditional RDD. For example, Anzia and Berry (2011) take a similar approach in their analysis of whether female members of Congress are more effective than their male counterparts at securing federal funds for their districts. Although Anzia and Berry's primary identification strategy is a differences-in-differences design using all congressional districts, for additional causal leverage the authors also estimate models using a sample that only includes districts with close elections (data limitations prevent Anzia and Berry from implementing a traditional RDD). We believe, however, that we are the first to take this approach using a simulated probability of receiving a treatment of interest.

3 An Application to Budget Deficits

We demonstrate both of our approaches with an application to state budgeting. Specifically, we reevaluate the hypothesized link between divided government and budget deficits/surpluses. During the 1980s and early 1990s observers of American politics frequently blamed the country's rising national debt on split partisan control of government. This argument was most clearly articulated by McCubbins (1991) who observed that, following the 1981 tax cut, a Democrat-controlled House of Representatives and Republican-controlled presidency were stuck in a fiscal stalemate. Both wanted to increase spending on their preferred priorities—Democrats wanted to spend more on domestic programs and Republicans more on defense—while paying for these increases by cutting expenditures elsewhere. According to McCubbins, the impasse was resolved by increasing spending in areas that each party preferred, without making cuts elsewhere. Similarly, there was no agreement on revenue increases to pay for these expenditures, leading to runaway budget deficits.

While McCubbins' analysis is deeply grounded in the politics of the 1980s, other scholars make similar though less temporally-specific arguments. The underlying logic is typically that the production of a balanced budget requires cooperation and coordination across branches, which should be more difficult during periods of split partisan control (Roubini and Sachs 1989a, 1989b; Cox and McCubbins 1991; Alesina and Perotti 1994; Alt and Lowry 1994; Poterba 1994; Hahm, Kamlet, and Mowry 1997; Krouse 2000).

Despite the appeal of these arguments, the empirical evidence on the fiscal effects of divided government is mixed. Though the 1980s certainly saw large increases in deficit spending, other periods of divided government (most notably the late 1990s) produced fiscal balance. Indeed, while several studies uncover evidence indicating that divided government increases the size of deficits (cf., McCubbins 1991; Cox and McCubbins 1991; Roubini and Sachs 1989a, 1989b), others either do not (cf., Hahm, Kamlet, and Mowry 1997) or conclude that the link is much more nuanced. For example, Poterba (1994) and Alt and Lowry (1994) both find that while divided government does not necessarily lead to increased deficits, it lengthens the amount of time it takes government to respond to both negative and positive fiscal shocks. Krause (2000), on the other hand, concludes that it is the amount of ideological policy divergence across political institutions, rather than divided government itself, that affects deficit size.

The unresolved nature of this debate, combined with the large amount of available data on state budgeting, creates a useful arena for demonstrating the approaches that we develop here.

One caveat of studying state budget deficits is that deficits tend to occur with less frequency and to be smaller at the state level than deficits at the national level. This is due, in part, to the fact that all states (with the exception of Vermont) have some form of balanced budget requirement. Of these 49 states, 36 have the particularly stringent requirement of a "no carry-over rule" which forbids deficits from being carried forward from one fiscal period into the next. That being said, state balanced budget requirements lack real enforcement mechanisms. The presence of these requirements, though, may mean that divided government will have a smaller effect on the size of deficits than it would at the national level (i.e., the state-level represents a hard test of the hypothesis).

For data on state spending and revenue we access the U.S. Census Bureau's Data Base on Historical Finances of State Governments, which includes detailed information about state fiscal policy in all 50 states from 1942-2008. Though our main interest is budget deficits, we opt to use a measure of budget surplus as our key dependent variable. We do this for ease of interpretation. Surplus is operationalized as the change in the difference between general revenue and general expenditure from one year to the next (measured in per capita constant dollars). Using this operationalization allows us to avoid the potential confusion that may arise from making reference to negative deficits. Because we uncover evidence of the anticipated divided government effect, we also estimate additional models to understand whether this result is driven by increases in spending, decreases in revenue, or some combination of both. Changes in spending and revenue are also measured in per capita constant dollars.[9] All of our analyses omit both Alaska and Wyoming, which tend to have large fluctuations in budget surpluses that are driven largely by changes in the prices of natural resources. Summary statistics for our fiscal variables are presented in Table 1.

4 Results

Below we consider the effect of divided government on the size of state budget surpluses. We first utilize our distance to divided government forcing variable within a traditional sharp RDD

[9]Our results are substantively similar if we operationalize these dependent variables as percent change or change as a share of state personal income.

framework; then we replicate our analyses utilizing our simulated probability of divided government measure and the PRD framework.

4.1 Distance to Divided Government Forcing Variable

Before presenting our RDD results, we begin by conducting the familiar set of tests to insure that the key identification assumption holds, that is, that potential outcomes are smooth across the discontinuity in the forcing variable. We suspect that the "no sorting assumption" will be easily met since our forcing variable is composed of electoral results for multiple offices, making precise control over the forcing variable implausible. That being said, we still evaluate the validity of our design in several ways. We begin with the McCrary (2008) test to assess the density of the forcing variable at the cutpoint. As expected, we fail to reject the null hypothesis of no sorting.[10] Next, we conduct a series of placebo tests and check for imbalances in baseline covariates of observations that are near the threshold but differ in treatment assignment. The covariates we consider are: the presence of stringent balanced budget requirement (i.e., a non-carry over rule), the use of biennial budgeting, legislative session length, the presence of a supermajority requirement for budget adoption, whether the governor possesses line item veto power, whether the budget is adopted during an election year, and the change in per capita income from the previous year. The results of these placebo tests (presented in Table 2) provide support for the validity of our design.

We next present the standard RDD plot (Figure 8) with our forcing variable (distance to divided government) on the x-axis and our outcome of interest (change in the budget surplus) on the y-axis. Using this plot we look for graphical evidence of a change (either a jump or dip) in the value of the outcome variable at the threshold. In other words, we see whether going from barely having unified government to barely having divided government leads to an observable change in the size of state's surplus. Observing a change at the cutpoint would provide preliminary evidence of a causal relationship. Lines (4th order polynomials) on either side of the cutpoint plot the relationship between the change in surplus and the distance to divided government.[11] At the

[10] The log difference in heights 0.15 with standard error 0.213 ($p = 0.482$).

[11] We follow common RDD guidance to plot flexible polynomials on either side of the threshold in the forcing variable (Imbens and Lemieux 2008; Lee and Lemieux 2010). This approach allows us to graphically assess any discontinuity at the cutpoint while mitigating the concern that we could mistake a nonlinear relationship for a discon-

cutpoint, we do observe a dip in the change in surplus from about \$60 to -\$10 per capita as a state moves from unified to divided government.[12] For some states, a decrease in the per capita change in surplus could simply shrink a surplus, but for others it could increase the size of the deficit or even shift a state from having a surplus to facing a deficit.

Following this, we estimate regression models to more rigorously test for the presence of a divided government effect. In keeping with current best practices we employ local linear models using only those observations that lie within a specified bandwidth on either side of the cutpoint (Gelman and Imbens 2014; Skovron and Titiunik 2015; cf. Imbens and Lemieux 2008). Observations within the bandwidth are weighted based on their proximity to the threshold. Specifying a bandwidth is not entirely straightforward because it involves a tradeoff between bias and variance. Wider bandwidths can lead to biased estimates because they incorporate observations further from the discontinuity. However, particularly narrow bandwidths can produce unbiased estimates though a smaller number of observations generally increases the variance of the estimates. Although electoral RDDs commonly use a bandwidth of 5%, researchers do not necessarily agree on a single approach. Because the choice of bandwidth can alter RDD results, some advocate relying on data-driven techniques to minimize researchers' discretion (Imbens and Kalyanaraman 2012; Skovron and Titiunik 2015). We follow this advice and employ the optimal bandwidth calculated using the algorithm recommended by Calonico, Cattaneo, and Titiunuk (2014), hereafter referred to as the CCT bandwidth. We present these results as well as results estimated with the commonly-utilized bandwidths of 5% and 10%. To assess the sensitivity of our results to bandwidth size, we also replicate our analyses across bandwidths ranging from 0.03 to 0.20, in increments of 0.01.

Our results, presented in Table 3, provide further evidence that divided government does have a negative effect on the change in budget surplus. The CCT optimal bandwidth of 5.1% (Column 3) and a bandwidth of 5% (Column 1) produce nearly equivalent estimates indicating

tinuity.

[12]While Figure 8 clearly shows a discontinuity at the threshold in our forcing variable, the trends farther from the threshold are noisier. Unfortunately, we cannot provide an explanation for these patterns. The trade off for causal identification in an RDD is that the results are local. Indeed, best practices for RDD require focusing only on observations that lie close to the cutpoint (Skovron and Titiunik 2015). This focus on close observations is how an RDD overcomes the threat of endogeneity.

that divided government leads to a \$107 decrease in per capita budget surplus. These results are large in magnitude and statistically significant. Using a 10% bandwidth (Column 2), we estimate an effect that is smaller in magnitude (\$53) and falls just short of conventional levels of statistical significance ($p = 0.12$). Figure 9 plots the coefficients from local linear models across a variety of bandwidths. Overall, it suggests that our results are quite stable across the narrower bandwidths where states are most similar in their likelihood of experiencing divided government. These results also are substantively quite striking. For states with a surplus less than \$107 per capita (keep in mind the mean value for surplus is \$45 per capita), moving from unified to divided government would, on average, lead to a budget deficit.

To unpack this result further, we estimate RDD models in which the dependent variable is the per capita change in either expenditures or revenues. This allows us to understand more fully the channels through which divided government affects budget deficits. To save space, we only present a coefficient plot of these results (see Figure 10). For each new dependent variable we present two point estimates, one using a 5% bandwidth and the other using the CCT optimal bandwidth (note that the CCT optimal bandwidth changes somewhat with each dependent variable). We estimate the effect of divided government on the change in per capita general expenditures is an increase of about \$74, using either bandwidth. This effect is both substantively and statistically significant. Interestingly, our estimates of the effect of split partisan control on revenues are negative and relatively small, failing to reach conventional levels of statistical significance using either bandwidth. In combination, these results suggest that the larger deficits/smaller surpluses that result from divided government are driven by increased expenditures without corresponding increases in revenues. This is consistent with the model of budgeting under divided government that McCubbins (1991) used to explain the rising federal deficit in the 1980s and early 1990s.

Though we do not report the results here, we considered whether rules prohibiting states from carrying over deficits from one year to the next might condition the effects of divided government on surplus size. We find little, if any, evidence that such restrictions matter. However, it is difficult to analyze heterogeneous treatment effects because most states have these stringent balanced budget requirements. The relatively small amount of cross-state variation is compounded by the fact that, by the nature of an RDD, we are estimating effects using a relatively small sample size. If we include an indicator for a no-carryover rule in our RDD specifications, our results remain unchanged and the coefficient of no-carryover rule is small and statistically indistinguishable from zero. We also include these rules as a covariate when we validate our RDD and find no evidence of a discontinuity at the threshold of distance to divided government, which suggests that a state with divided government is no more (or less) likely to have a stringent balanced budget rule than a state with unified government (see Table 2).

4.2 Simulated Probability of Divided Government

We replicate the above analyses using our PRD approach. Recall, this approach utilizes our simulated probability of divided government measure to restrict our sample to those observations for which the odds of experiencing divided government are close to 50-50. With this sample, we then estimate standard unweighted OLS regression models. One challenge, is that for a probability measure there is not a consensus as to the range of values that would constitute a close election, and because there is no clear threshold, existing methods for calculating an optimal bandwidth cannot be used. As a result we simply opt to present full regression estimates using the 5% and 10% bandwidths that are commonly employed in sharp RDD analyses, though at these bandwidths we have many fewer observations than in the analyses that employed the distance to divided government forcing variable. To evaluate the sensitivity of our results to bandwidth size, we again replicate our analyses across bandwidths ranging from 0.03 to 0.20, in increments of 0.01.

Our first regression results are shown in Table 4. These are very similar to what we previously observed. At a 5% bandwidth, we find that divided government leads to a $64 decrease in the size of the surplus, and at the 10% bandwidth a $51 decrease. Though these effect sizes are a bit smaller than our traditional RDD estimates (and somewhat noisier), these results remain both substantively and statistically meaningful. Figure 11 shows that these results are robust to bandwidth size, though once we get to bandwidth sizes of around 15% the effect size shrinks and is no longer statistically significant (at least at conventionally employed levels of significance).

However, by the time we approach these larger bandwidths, it becomes much less plausible that we are comparing observations where treatment assignment is as if random.

When we consider the effect of divided government on expenditures and revenues, we again find a pattern that mirrors the expectations of McCubbins (1991). Figure 12 shows that divided government leads to an increase in expenditures but has no distinguishable effect on revenue. It is worth noting that these results generated using our modified RDD approach are, again, unchanged if we include a measure of balanced budget stringency in OLS models. Likewise, we find no evidence that effects are heterogeneous across states, at least with respect to rules that limit a state's ability to carry a deficit from one fiscal year or biennium to the next.

4.3 More Traditional Approaches

Like all RDDs, our approaches account for selection into treatment—here, the likelihood a state will experience divided government. The skeptical reader, however, might ask what happens if we do not. To explore this possibility, we estimate four models that rely on neither of the identification strategies that we develop above.

The results of these estimations are reported in Table 5. The first model is a simple bivariate regression that regresses the change in surplus on an indicator variable for divided government. The second model includes a handful of additional predictors: the size of government (operationalized as total expenditures per capita), a indicator for the use of biennial budgeting, the presence of a no-carryover rule[13], and change in income (operationalized as the annual percent change in real personal income per capita). The third and fourth models add state and year and fixed effects to the prior model.

None of these models find anything close to a statistically meaningful divided government effect. Though we do not report the results here, we also find noisy null results if we use either per capita expenditures or per capita revenues as the dependent variable.

We view the results presented in Table 5 not as evidence that divided government is inconsequential but as evidence of the difficulty of testing for an effect using traditional methods.

[13] A no-carryover rule is the most stringent type of balanced budget requirement, which prohibits states from transferring budget deficits from one year to the next.

It is only when we account for the underlying probability of divided government that we uncover robust evidence of the anticipated effect. Indeed, these results further demonstrate the usefulness of the approaches for causal identification that we develop above. While it may be possible to find similar effects using traditional methods (i.e., by estimating regression models that state and year fixed effects plus covariates not utilized here) it remains unclear which variables ought to be included in models or even whether all necessary variables can be measured.[14]

5 Discussion

Conventional wisdom views divided government as shaping a variety of outcomes, including legislative productivity, the size of government, the timeliness of legislative action, the size of budget deficits, etc. Empirically evaluating these claims has proven to be tricky because the presence of divided government is not randomly distributed. One increasingly common strategy for addressing concerns of endogeneity with observational data is the use of regression discontinuity designs. RDDs, however, are difficult to apply to treatments like divided government, because the presence (or absence) of divided government is the result of elections to multiple institutions, making the construction of a forcing variable problematic.

We present two approaches for dealing with this challenge. We begin with a series of simulations in which we apply shocks of varying sizes to real-world state-level election results. The results of these simulations can be used to produce measures that identify state years in which unified and divided partisan control of government are almost equally as likely. The first such measure—distance to divided government—captures the smallest state-level vote shock that would result in a different outcome than the one actually observed. Because this measure has a sharp discontinuity between unified and divided government, it can be used within the familiar sharp RDD framework. The second measure—the probability of divided government—is operationalized as the proportion of simulations in which neither party wins full control of government. Because our probability measure lacks a threshold where treatment assignment changes discontinuously, it is

[14]This uncertainty, accompanied by researchers' discretion over regression models, also raises concerns about the potential for data mining and endless debates about proper specification.

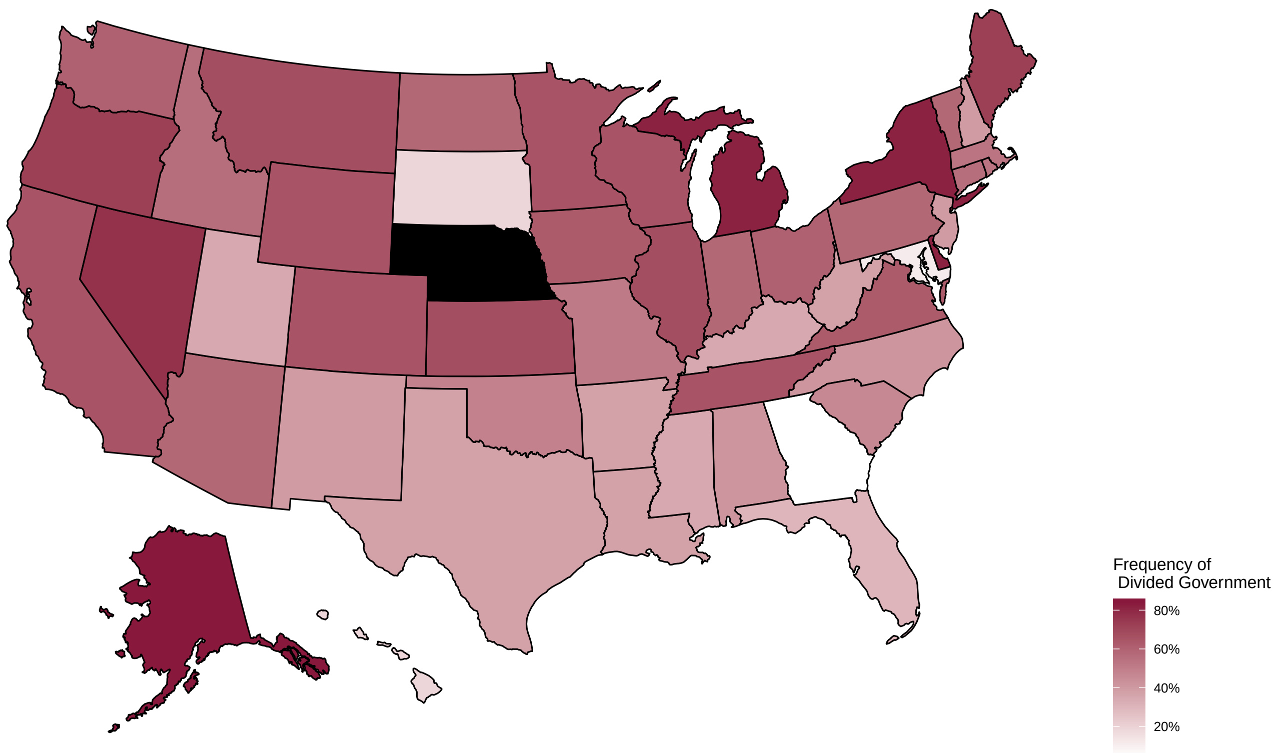
not a suitable forcing variable for an RDD. We rely on the intuition behind the RDD, however, and use traditional OLS regression to estimate the effect of divided government focusing on the subset of observations where the odds of experiencing divided government are close to 50-50 (PRD). Both of the approaches we present here produce a causal estimate of the local average treatment effect of divided government.

We demonstrate an application of these approaches by utilizing them to reevaluate the hypothesis that divided government brings about larger budget deficits. Regardless of whether we employ our distance to divided government measure in combination with a sharp RDD or employ the probability of divided government measure in combination with a PRD, we find evidence of a causal relationship. The average size of this effect is large enough to take a state from fiscal surplus to fiscal deficit. The increase in budget deficits during periods of divided government appears to be driven by increases in spending that are not accompanied by corresponding increases in revenue.

Though both of our empirical strategies produce similar causal estimates of the effect of divided government, researchers may prefer one approach to the other (we, however, are agnostic on the matter). The first approach enables researchers to use the familiar and widely-accepted RDD toolkit, which includes a set of established best practices for analyzing data and validating the design and its identification assumptions. One potential concern with this approach (as well as other electoral RDDs), however, is that while the forcing variable tells us how large of an electoral swing would be needed to produce a different outcome, it does not address how likely that electoral swing is to occur. As our analysis of historical election data demonstrates, for example, a shock of 4% is more likely in some states than other states. Our second approach takes these differences into account (and as a result identifies fewer observations where the odds of experiencing divided government are very close to 50-50). That being said, this approach is not yet an established strategy for causal identification, and probability measures may be somewhat sensitive to features of the simulations, such as the distribution from which historical electoral shocks are drawn.

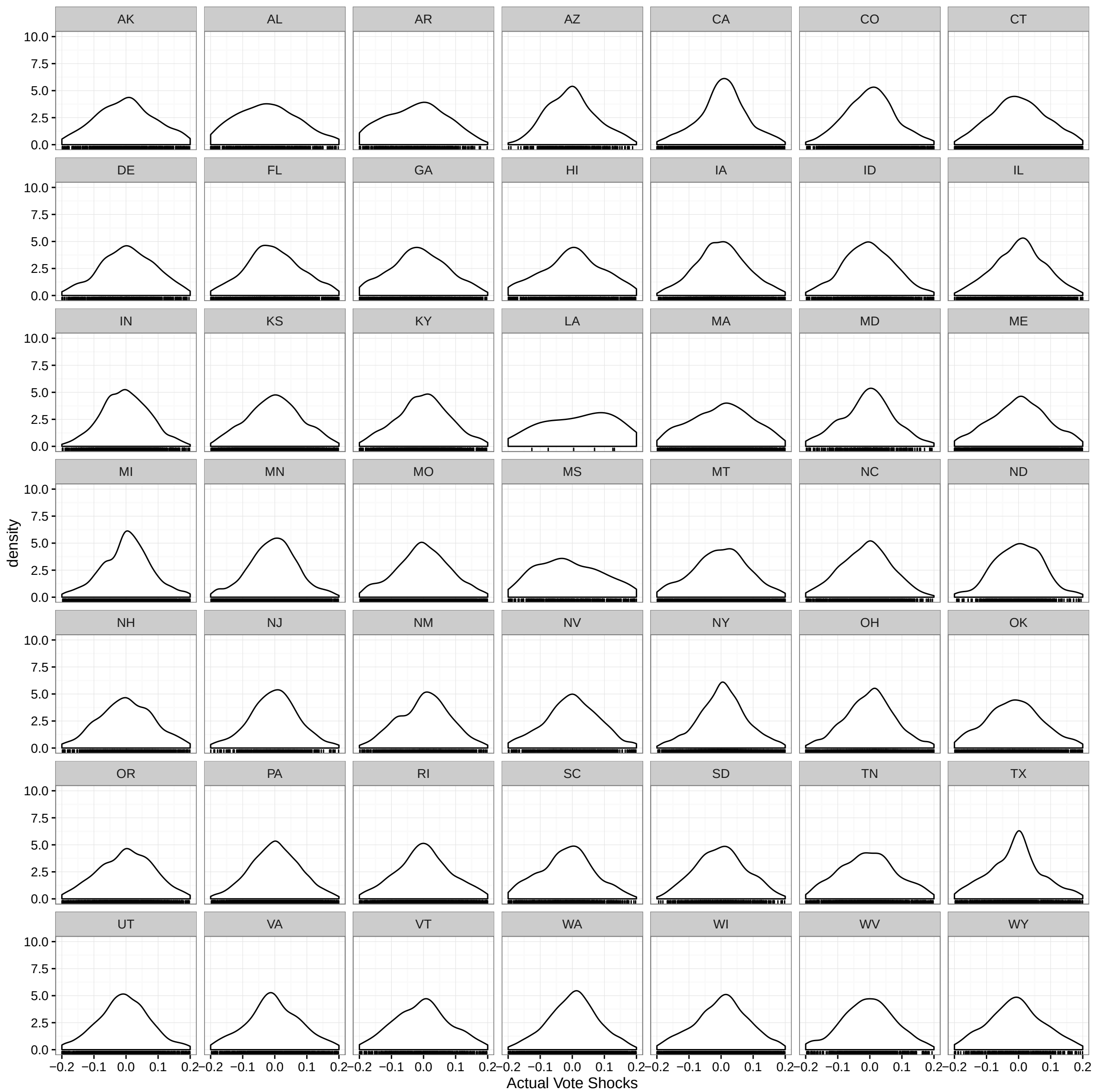
Overall we believe that both of these approaches, by expanding the causal inference toolkit available to researchers, can be used to reinvigorate empirical inquiry into the consequences of divided government. Long debated questions such as the effect of divided partisan control on legislative productivity can be reevaluated with new rigor. Our simulations can also be improved upon by potentially incorporating other factors that are known to shape electoral outcomes including the partisanship of the electorate, presidential approval, the state of the economy, etc. Moving forward, our simulated probability measure, in particular, offers the potential to move beyond local average treatment effects by using a design that incorporates weighting (e.g., inverse probability weighting, propensity score weighting) to plausibly identify a global average treatment effect. Finally, beyond the study of divided government, scholars can also adapt the approaches we develop here to causally explore other questions that are central to political science, such as the impact of partisan control of government.

Figure 1: Divided Government



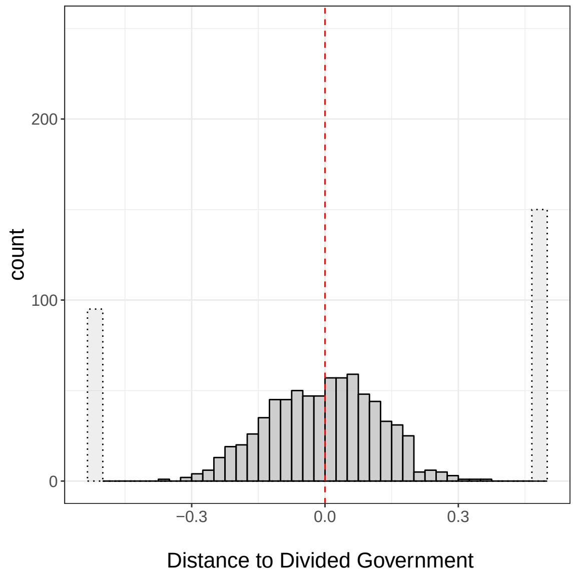
Note: Frequency of divided government by state, 1968-2010. Nebraska is excluded because of its nonpartisan legislature.

Figure 2: Historical Election Shocks



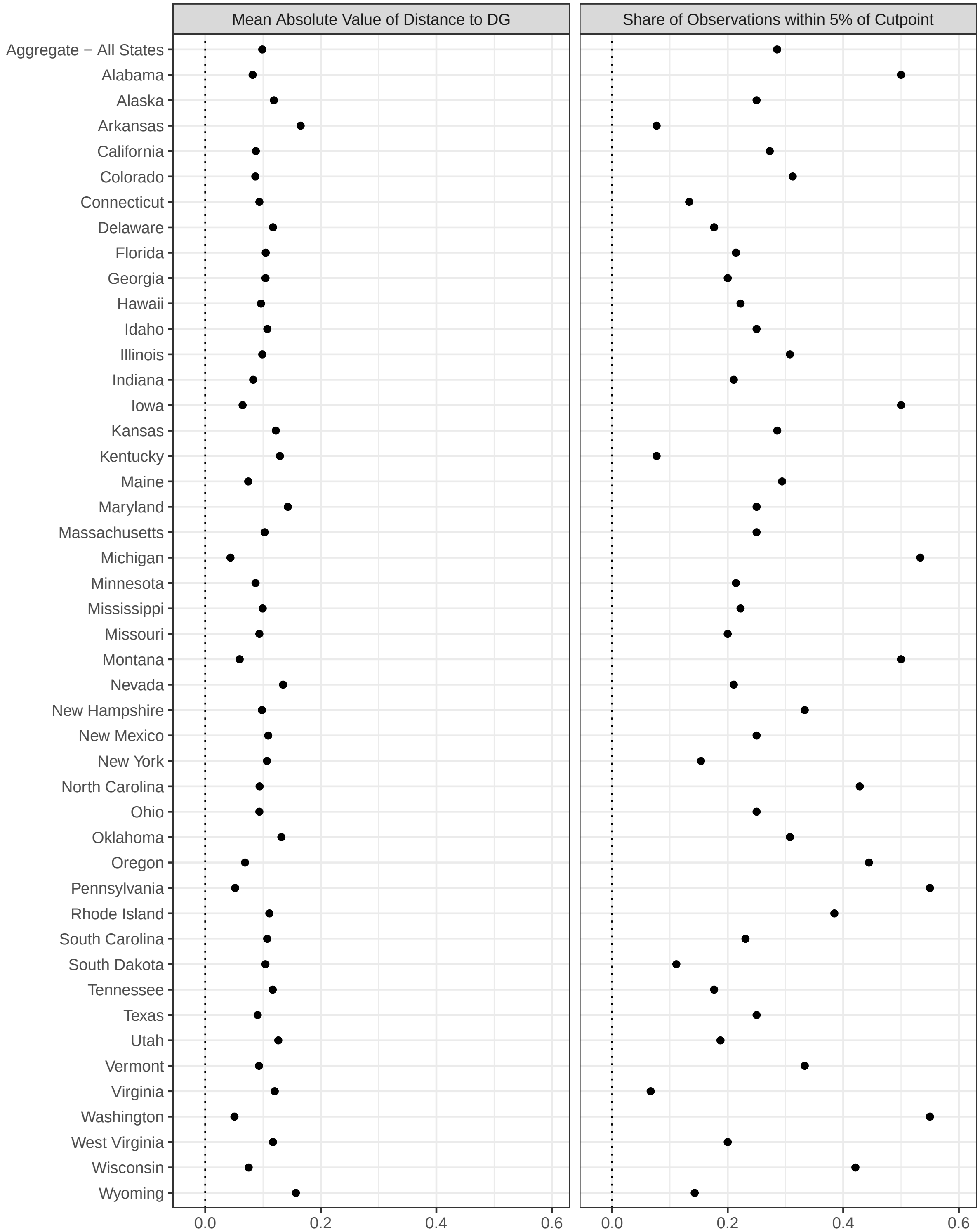
Note: The x-axis for each state indicates the size of vote shocks. The figures plot, by state, the density of historical vote shocks, including shocks to state assembly, state senate, and gubernatorial elections.

Figure 3: Histogram of Forcing Variable



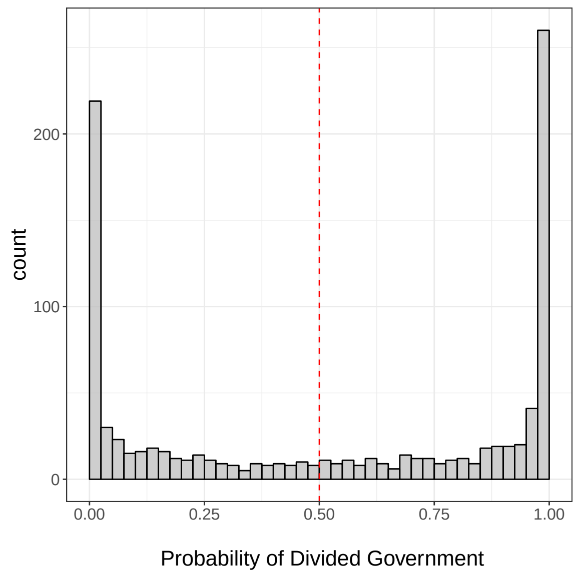
Note: The histogram displays the distribution of the forcing variable. Zero on the x-axis is the cutpoint. Observations to the right of the cutpoint (i.e., positive values) have divided government; observations to the left of the cutpoint (i.e., negative values) have unified government. The y-axis is a count of the number of state years that fall into each bin. Bins (shaded light grey with dotted outlines) at either end of the distribution denote observations with missing values.

Figure 4: Distance to Divided Government



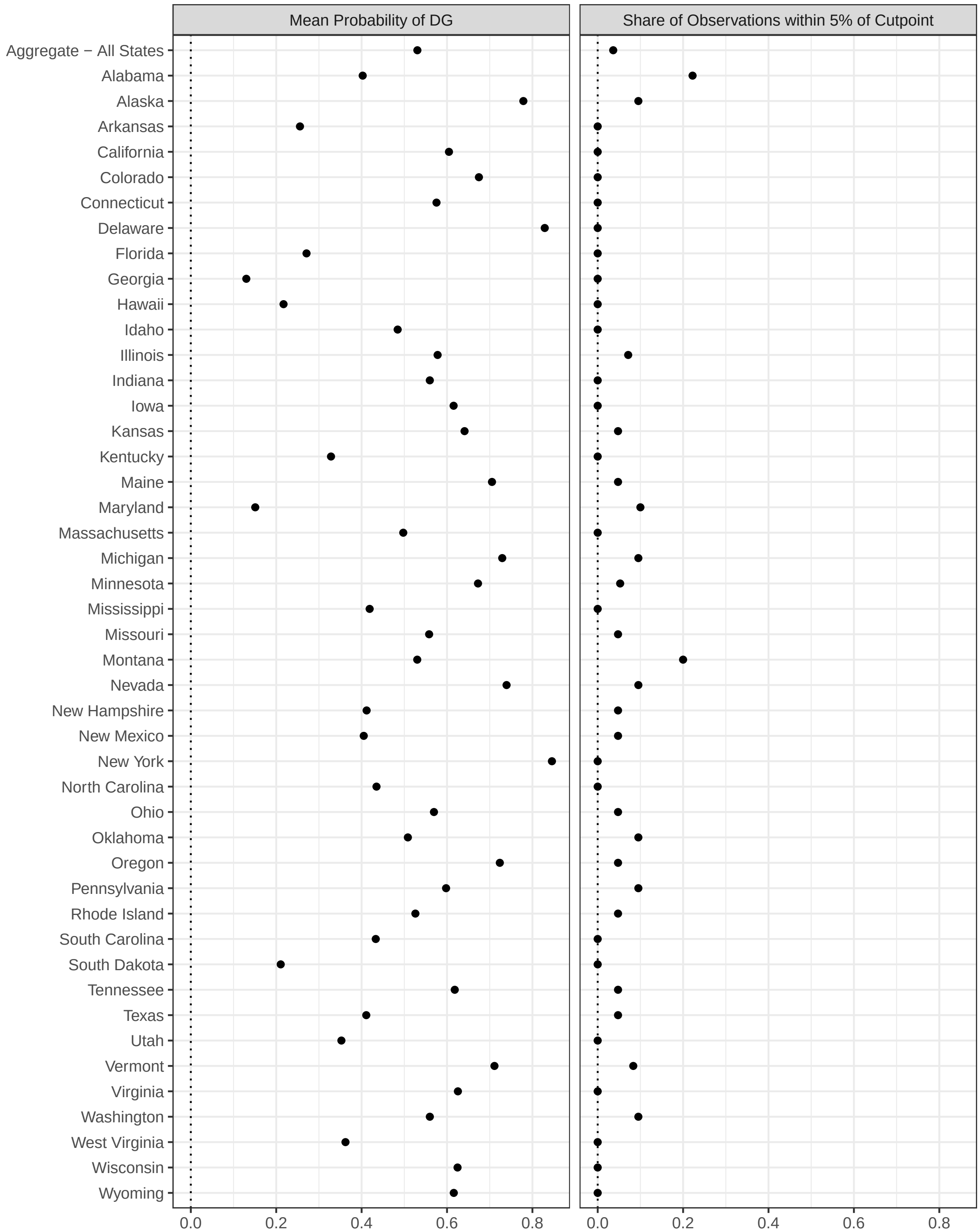
Note: The panel on the left shows the mean of the absolute distance to divided government on the x-axis for each state on the y-axis. The right panel indicates the share of state observations on the x-axis for which the absolute value of the distance to divided government forcing variable is within ± 5% of the cutpoint at 0.

Figure 5: Histogram of Forcing Variable



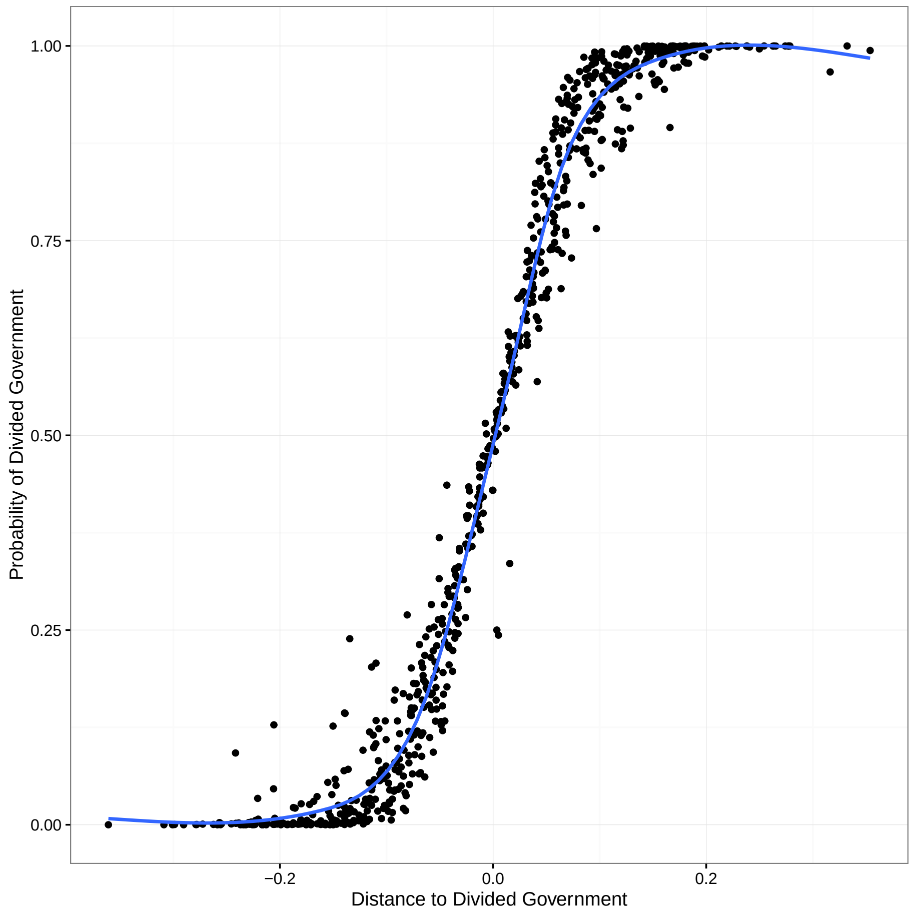
Note: The histogram displays the distribution of the simulated probability of divided government. The x-axis is the probability of divided government, and the y-axis is a count of the number of state years that fall into each bin.

Figure 6: Probability of Divided Government



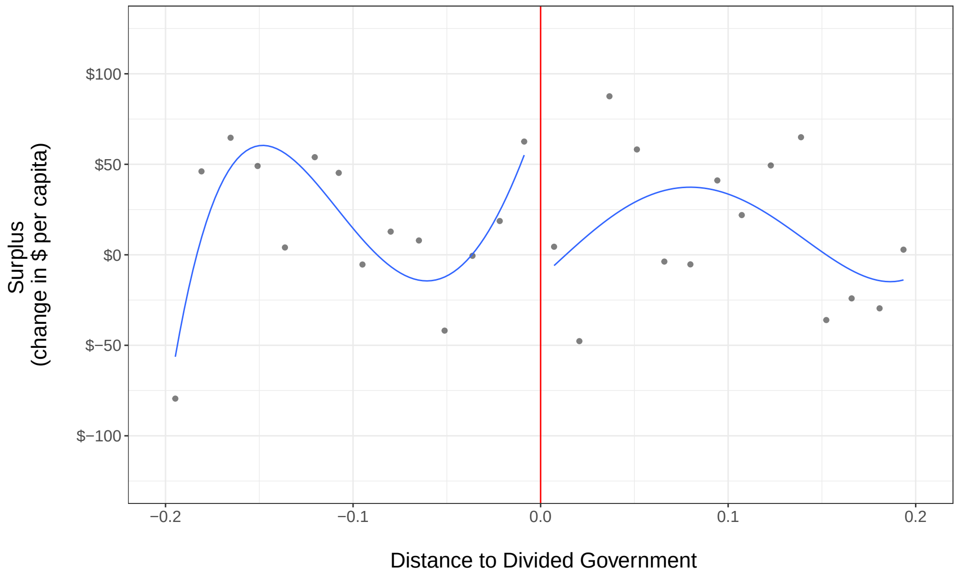
Note: The panel on the left shows the mean of the simulated probability of divided government on the x-axis for each state on the y-axis. The right panel indicates the share of state observations on the x-axis for which the the probability of divided government forcing variable is within ± 5% of 50%.

Figure 7: Distance vs. Probability



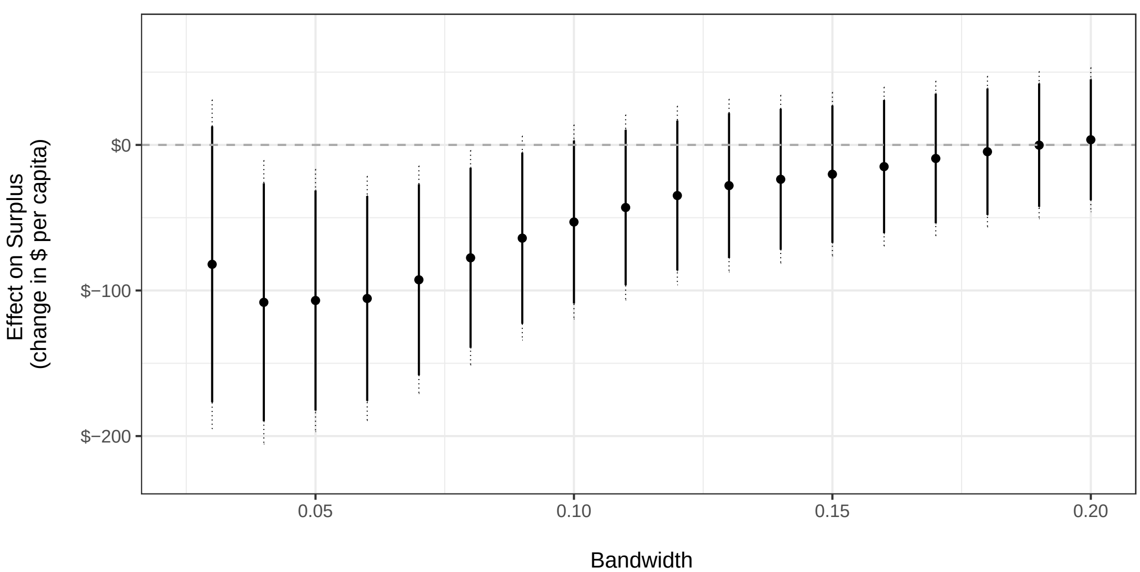
Note: The x-axis is the distance to divided government centered at 0, and the y-axis is the probability of divided government. Both measures were generated using our simulations-based method.

Figure 8: Surplus & Divided Government



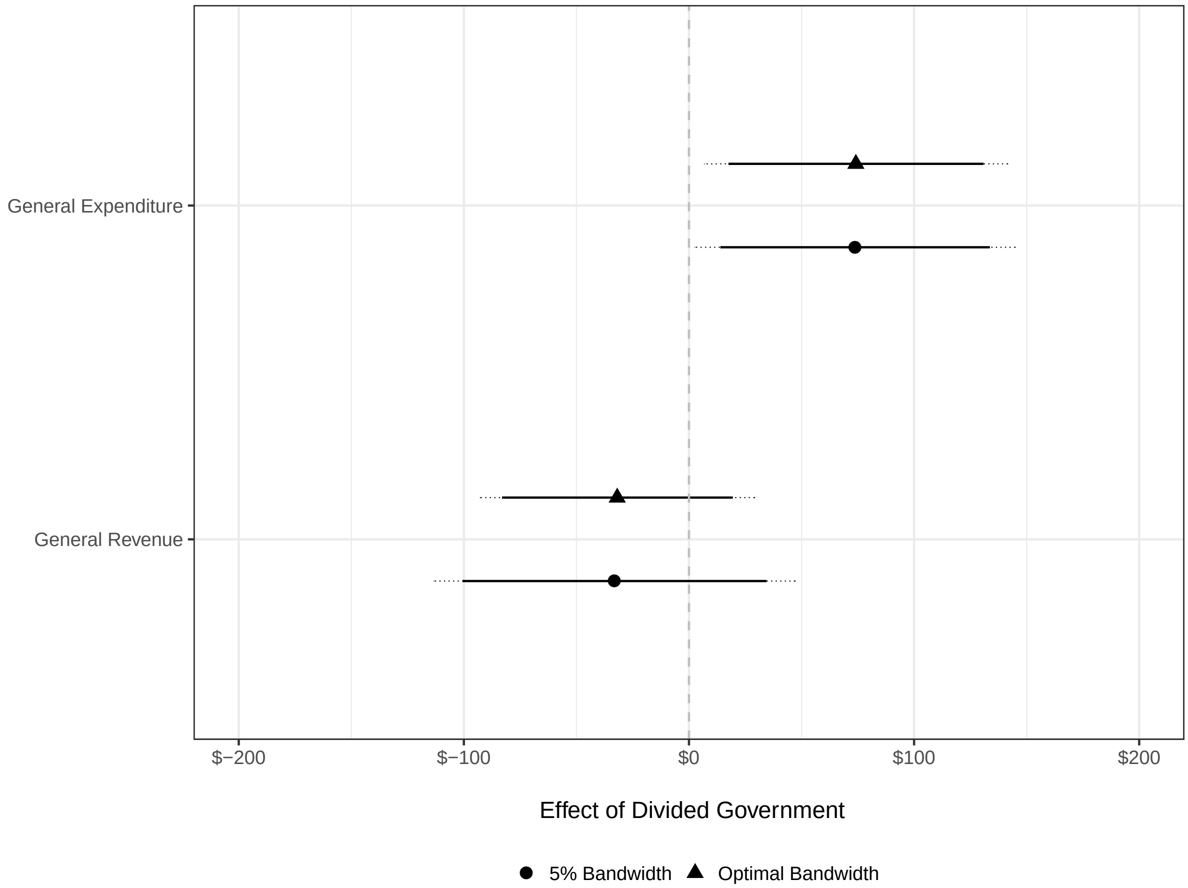
Note: The x-axis is the distance to divided government centered at 0, and the y-axis is the change in the budget surplus (measured in per capita dollars). The points are averages of the change in surplus within 1.5% bins.

Figure 9: Surplus & Divided Government



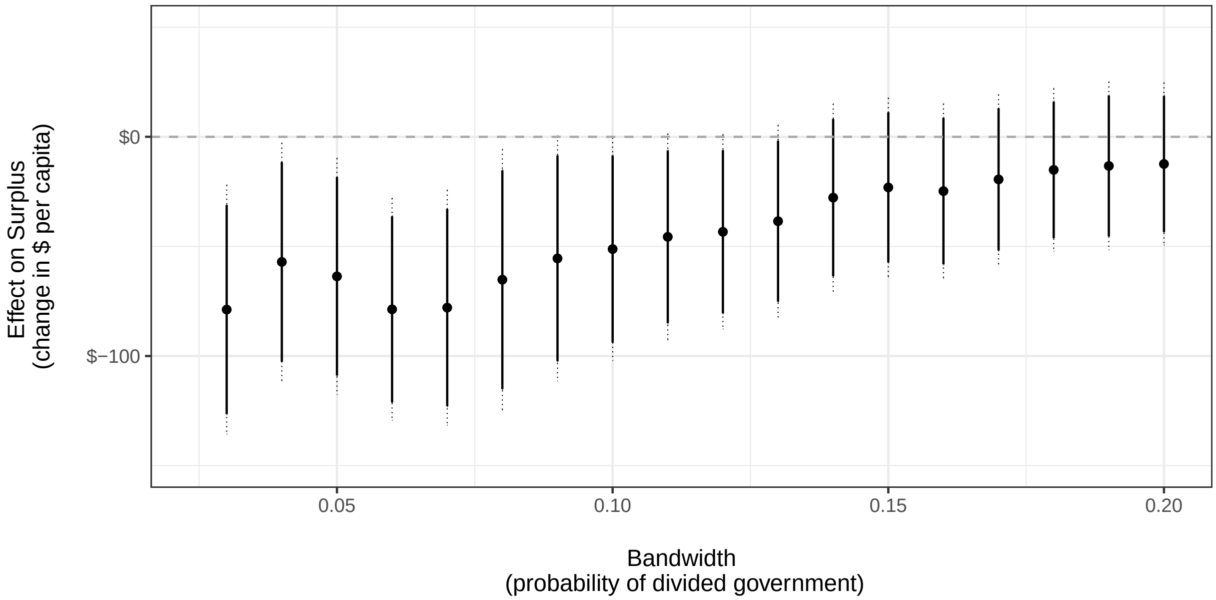
Note: The y-axis measures the effect size while the x-axis indicates the bandwidth. Dots indicate point estimates from local linear regression models, and the error bars reflect two-tailed tests. The solid black lines show 90% confidence intervals while the dotted lines indicate 95% confidence intervals.

Figure 10: Effect of Divided Government



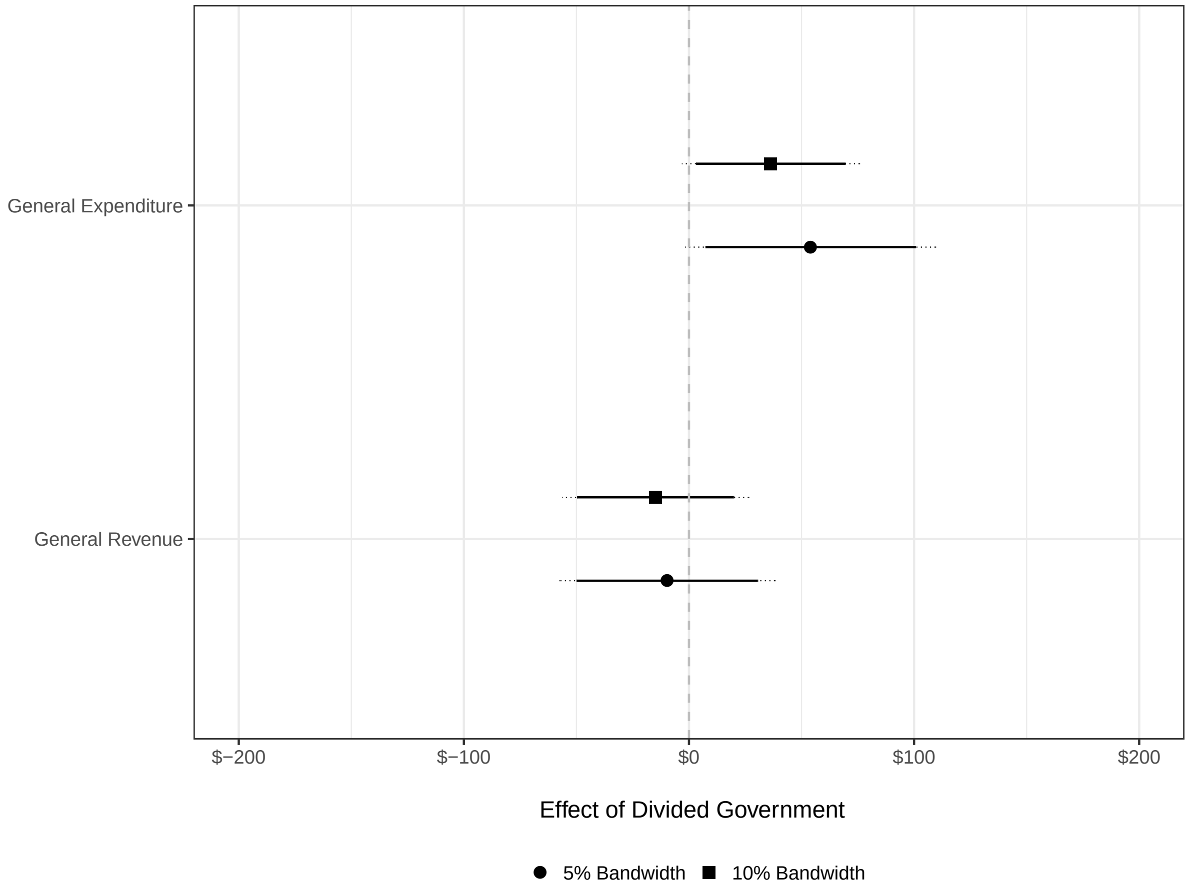
Note: The x-axis measures the effect of divided government on the per capita change in the dependent variables on the y-axis. Dots indicate point estimates from local linear regression models using a 5% bandwidth, and triangles indicate point estimates from similar specifications using the CCT optimal bandwidth (for general expenditure, 0.056; for general revenue, 0.088). The error bars reflect two-tailed tests with solid black lines showing 90% confidence intervals and dotted lines indicating 95% confidence intervals.

Figure 11: Surplus & Divided Government



Note: The y-axis measures the effect size while the x-axis indicates the bandwidth. Dots indicate point estimates from local linear regression models, and the error bars reflect two-tailed tests. The solid black lines show 90% confidence intervals while the dotted lines indicate 95% confidence intervals.

Figure 12: Effect of Divided Government



Note: The x-axis measures the effect of divided government on the per capita change in the dependent variables on the y-axis. Dots indicate point estimates from OLS regression models that incorporate observations within ± 5% of 50% probability of divided government. Squares indicate point estimates from similar specifications with the sample restricted to observations within ± 10% of 50%. The error bars reflect two-tailed tests with solid black lines showing 90% confidence intervals and dotted lines indicating 95% confidence intervals.

Table 1: Summary Statistics: Dependent Variables

Outcome ($ per capita)	Mean	Minimum	Maximum
Change in Surplus	2.22	-678.83	677.05
Change in Expenditure	62.93	-512.31	896.81
Change in Revenue	65.14	-484.49	794.86

Note: Reported in constant dollars based on data from the U.S. Census Bureau's Data Base of Historical Finances of State Governments.

Table 2: RDD Validity Tests

Covariate	Estimate	Std. Error	p-value	Bandwidth
	5% Bandwidth			
Lagged Change in Expenditure	-23.413	34.245	0.495	0.050
Lagged Change in Surplus	-26.401	29.583	0.373	0.050
Lagged Change in Revenue	-49.814	34.693	0.153	0.050
No-Carryover Rule	0.103	0.146	0.480	0.050
Biennial Budgeting	0.053	0.161	0.742	0.050
Session Length	23.762	23.555	0.316	0.050
Supermajority Budget Rule	-0.022	0.051	0.660	0.050
Per-capita Income Change	1.129	1.026	0.273	0.050
Election Year	0.011	0.080	0.887	0.050
Line-item Veto	0.091	0.110	0.408	0.050
	Optimal Bandwidth			
Lagged Change in Expenditure	-10.684	25.189	0.672	0.103
Lagged Change in Surplus	-16.134	26.280	0.540	0.067
Lagged Change in Revenue	-28.470	29.438	0.334	0.076
No-Carryover Rule	0.066	0.103	0.522	0.102
Biennial Budgeting	0.019	0.121	0.878	0.096
Session Length	9.215	18.228	0.614	0.082
Supermajority Budget Rule	-0.022	0.049	0.657	0.067
Per-capita Income Change	0.802	0.818	0.328	0.077
Election Year	-0.006	0.061	0.916	0.081
Line-item Veto	0.032	0.091	0.727	0.090

Note: Estimates of local linear regression models. Maximum of conventional and robust standard errors reported. P-values reflect two-tailed tests.

Table 3: Surplus & Divided Government
Simulated Distance to Divided Government (RDD)

	Dependent Variable: Change in Surplus (per-capita $)		
	(1)	(2)	(3)
Divided government	-106.857^{**}	-52.946	-106.796^{**}
	(45.950)	(34.000)	(45.538)
Distance to divided gov't	$2{,}546.832^{*}$	$1{,}124.316^{**}$	$2{,}511.210^{**}$
	(1,306.704)	(546.373)	(1,274.237)
Distance to divided gov't * Divided government	-577.561	-504.842	-512.219
	(1,705.896)	(670.254)	(1,654.850)
Constant	82.124^{**}	55.749^{**}	81.695^{**}
	(36.983)	(27.627)	(36.713)
Bandwidth	0.050	0.100	0.051
Observations within Bandwidth	176	351	181
Residual Std. Error	97.744 (df = 172)	97.118 (df = 347)	97.722 (df = 177)
F Statistic	3.012^{**} (df = 3; 172)	2.701^{**} (df = 3; 347)	3.115^{**} (df = 3; 177)

Note: Estimates of local linear regression models. Maximum of conventional and robust standard errors reported. $^{*}p < 0.1$; $^{**}p < 0.05$ (two-tailed test).

Table 4: Surplus & Divided Government
Simulated Probability of Divided Government (PRD)

	Dependent Variable: Change in General Surplus (per-capita $)	
	(1)	(2)
Divided government	-63.661^{**}	-51.197^{*}
	(27.709)	(25.915)
Constant	48.253^{**}	33.509^{*}
	(20.216)	(18.399)
Bandwidth	0.050	0.100
Observations within Bandwidth	62	125
Residual Std. Error	108.864 (df = 60)	143.701 (df = 121)
F Statistic	5.278^{**} (df = 1; 60)	3.903^{*} (df = 1; 121)

Note: Estimates of OLS regression models. Maximum of conventional and robust standard errors reported. $^{*}p < 0.1$; $^{**}p < 0.05$ (two-tailed test).

Table 5: Surplus & Divided Government
Simulated Probability of Divided Government

	Dependent Variable: Change in General Surplus (per-capita $)			
	(1)	(2)	(3)	(4)
Divided government	−1.263	−2.305	−0.949	−3.109
	(6.007)	(6.055)	(5.072)	(4.807)
Size of government		0.003		0.042**
		(0.004)		(0.009)
Biennial budgeting		4.077		4.183
		(6.211)		(13.293)
No carry-over rule		−3.737		23.210**
		(6.895)		(10.629)
Change in real income		8.106**		2.708**
		(1.325)		(1.289)
Constant	2.920	−15.315	25.284**	−58.686**
	(4.385)	(11.280)	(11.518)	(20.748)
Fixed effects?	No	No	State & Year	State & Year
Observations	1,880	1,880	1,880	1,880
R^2	0.00002	0.029	0.162	0.170
Adjusted R^2	−0.001	0.026	0.122	0.129
Residual Std. Error	129.984 (df = 1878)	128.228 (df = 1874)	121.776 (df = 1793)	121.272 (df = 1790)
F Statistic	0.044 (df = 1; 1878)	11.166** (df = 5; 1874)	4.032** (df = 86; 1793)	4.130** (df = 89; 1790)

Note: Estimates of OLS regression models. Robust standard errors reported. $^{*}p < 0.1$; $^{**}p < 0.05$ (two-tailed test).

References

Alesina, Alberto and Roberto Perotti. 1994. “ The Political Economy of Budget Deficits.” National Bureau of Economic Research. Working Paper No. 4637.

Alt, James E., and Robert C. Lowry. 1994. “ Divided Government, Fiscal Institutions, and Budget Deficits: Evidence from the States” *American Political Science Review* 88: 811-28.

Angrist, Joshua D., and Jorn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Anzia, Sarah F. and Christopher R. Berry. 2011. “The Jackie (and Jill) Robinson Effect: Why Do Congresswomen Outperform Congressmen?” *American Journal of Political Science* 55(3): 478-93.

Binder, Sarah A. 2003. *Stalemate: Causes and Consequences of Legislative Gridlock*. Washington D.C.: Brookings Institution Press.

Binder, Sarah A. 2015. “The Dysfunctional Congress.” *Annual Review of Political Science* 18(7): 1-17.

Binder, Sarah A., and Forrest Maltzman. 2002. “Senatorial Delay in Confirming Federal Judges, 1947-1998.” *American Journal of Political Science* 46(1): 190-99.

Bowling, Cynthia J. and Margaret R. Ferguson. 2001. “Divided Government, Interest Representation, and Policy Differences: Competing Explanations of Gridlock in the 50 States.” *Journal of Politics* 63(1): 182-206.

Broockman, David E. 2014. “Do Female Politicians Empower Women to Vote or Run for Office? A Regression Discontinuity Approach.” *Electoral Studies* 34: 190-204.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs.” *Econometrica* 82(6):22952326.

Chiou, Fang-Yi and Lawrence S. Rothenberg. 2003. “When Pivotal Politics Meets Partisan Politics.” *American Journal of Political Science* 47(3): 503-22.

Coleman, John J. 1999. “Unified Government, Divided Government, and Party Responsiveness.” *American Political Science Review* 93(04): 821-835.

Cox, Gary W. and Mathew D. McCubbins, 1991. “Divided Control of Fiscal Policy.” In *Parallel Politics: The Politics of Economic Policy in Japan and the United States,* ed. S. Kernell. Washington, D.C.: Brookings Institution.

Dubin, Michael J. 2007. *Party Affiliations in the State Legislatures: A Year by Year Summary, 1796-2006*. Jefferson, NC: McFarland.

Erikson, Robert S. and Rocío Titiunik. 2015. "Using Regression Discontinuity to Uncover the Personal Incumbency Advantage." *Quarterly Journal of Political Science* 10(1): 101-119.

Ferreira, Fernando and Joseph Gyourko. 2014. "Does Gender Matter for Political Leadership? The Case of U.S. Mayors." *Journal of Public Economics* 112: 24-39.

Fiva, Jon H., Olle Folke, and Rune J. Sørensen. 2013. "The Power of Parties: Evidence from Close Municipal Elections in Norway." *The Scandinavian Journal of Economics* 120(1): 3-30.

Folke, Olle. 2014. "Shades of Brown and Green: Party Effects in Proportional Election Systems. *Journal of the European Economic Association* 12(5): 1361-1395.

Gelman, Andrew and Guido Imbens. 2014. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." National Bureau of Economic Research.

Gerber, Elisabeth R. and Daniel J. Hopkins. 2011. "When Mayors Matter: Estimating the Impact of Mayoral Partisanship on City Policy." *American Journal of Political Science* 55(2): 326-339.

Gray, Virginia and David Lowery, 1995. "Interest Representation and Democratic Gridlock." *Legislative Studies Quarterly* 20(4): 531-552.

Gubernatorial General Elections, all states, 1968-2010 summary. (2003). *CQ Voting and Elections Collection (Web site)*. Washington, DC: CQ Press. Retrieved from http://library.cqpress.com/elections/avg2008-4us1.

Hahm, Sung Deuk, Mark S. Kamlet, and David C. Mowery. 1997. "PostWar Deficit Spending in the United States." *American Politics Quarterly* 29: 139-167.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69(1): 201-209.

Hall, Andrew B. 2015. "What Happens When Extremists Win Primaries?." *American Political Science Review* 109(1): 18-42.

Hicks, William D. 2015. "Partisan Competition and the Efficiency of Lawmaking in American State Legislatures, 1991-2009." *American Politics Research* 43(5): 743-770.

Hopkins, Daniel J. and McCabe, Katherine T. 2012. "After It's Too Late: Estimating the Policy Impacts of Black Mayoralties in U.S. Cities." *American Politics Research* 40(4): 665-700.

Howell, William G. 2003. *Power without Persuasion: The Politics of Direct Presidential Action.* Princeton, NJ: Princeton University Press.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* 79(3): 1-27.

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

Kirkland, Patricia A. 2017. "The Business of Being Mayor: Mayors and Fiscal Policy in U.S. Cities." Unpublished manuscript, Columbmia University.

Kirkland, Patricia A. and Justin H. Phillips. 2018. “Is Divided Government a Cause of Legislative Delay?” *Quarterly Journal of Political Science* 13(2): 173-206.

Klarner, Carl. 2018. State Legislative Election Returns, 1967-2016. Harvard Dataverse, V3, UNF:6:pV4h1CP/B8pHthjjQThTTw==. https://doi.org/10.7910/DVN/3WZFK9

Klarner, Carl, William Berry, Thomas Carsey, Malcolm Jewell, Richard Niemi, Lynda Powell, and James Snyder. State Legislative Election Returns (1967-2010). ICPSR34297-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-01-11. http://doi.org/10.3886/ICPSR34297.v1

Kousser, Thad and Justin H. Phillips. 2012. *The Power of American Governors: Winning on Budgets and Losing on Policy*. New York: Cambridge University Press.

Krehbiel, Keith. 1998. *Pivotal Politics*. Chicago: University of Chicago Press.

Kriner, Douglas and Liam Schwartz. 2008. “Divided Government and Congressional Investigations.” *Legislative Studies Quarterly* 33(2): 295322.

Krause, George A. 2000. “Partisan and Ideological Sources of Fiscal Deficits in the United States.” *American Journal of Political Science* 44(3): 54159.

Lee, David. S. 2008. “Randomized Experiments from Non-random Selection in US House Elections.” *Journal of Econometrics* 142(2), 675-697.

Lee, David S. and Thomas Lemieux. 2010. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature* 48]: 281-355.

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

Mayhew, David R. 1991. *Divided We Govern*. New Haven, CT: Yale University Press.

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

McCubbins, Mathew D., 1991. “Party Politics, Divided Government, and Budget Deficits.” In *Parallel Politics: The Politics of Economic Policy in Japan and the United States,* ed. S. Kernell. Washington, D.C.: Brookings Institution.

Moraski, Bryon J., and Charles R. Shipan. 1999. “The Politics of Supreme Court Nominations: A Theory of Institutional Constraints and Choices.” *American Journal of Political Science*, 43(4): 106995.

Nicholson, Stephen P., Gary M. Segura, and Nathan D. Woods. 2002. “Presidential Approval and the Mixed Blessings of Divided Government.” *American Journal of Political Science* 50(1): 14659.

Poterba, James. 1994. “State Response to Fiscal Crises: ‘Natural Experiments’ for Studying the Effects of Budgetary Institutions.” *Journal of Political Economy* 102: 793-821.

Rogers, James R. 2005. “The Impact of Divided Government on Legislative Production.” *Public Choice*. 123 (April): 217-233.

Rosenthal, Alan. 2004. *Heavy Lifting: The Job of the American Legislature*. Washington, D.C.: CQ Press.

Roubini, Nouriel, and Jeffrey D. Sachs. 1989a. “Political and Economic Determinations of Budget Deficits in the Industrial Democracies.” *European Economic Review*. 33: 903-33.

Roubini, Nouriel, and Jeffrey D. Sachs. 1989b. “Government Spending and Budget Deficits in the Industrial Countries.” *Economic Policy*. 8: 99-132.

Skovron, Christopher and Rocio Titiunik. 2015. “A Practical Guide to Regression Discontinuity Designs in Political Science.” Working paper. http://sites.lsa.umich.edu/webbkeane/wp-content/uploads/sites/233/2015/10/SkovronTitiunik-v5.pdf

Squire, Peverill. 1998. “Membership Turnover and the Efficient Processing of Legislation.” *Legislative Studies Quarterly* 23(1): 23-32.