Is Divided Government A Cause of Legislative Delay?

Patricia A. Kirkland
Center for the Study of Democratic Institutions
Vanderbilt University
patricia.a.kirkland@vanderbilt.edu

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

August 21, 2017

Abstract

Despite the compelling theoretical prediction that divided government decreases legislative performance, the empirical literature has struggled to identify a causal effect. We suspect that a combination of methodological challenges and data limitations are to blame. Here, we revisit this empirical relationship. Rather than relying on traditional measures of legislative productivity, however, we consider whether divided government affects the ability of lawmakers to meet critical deadlines—specifically, the ability of state lawmakers to adopt an on-time budget (as mandated by state law). By focusing on delay instead of productivity we avoid measurement problems, particularly the challenges inherent in measuring the supply of and demand for legislation. To assess the causal effect of divided government, we develop and implement a regression discontinuity design (RDD) that accounts for the multiple elections that produce unified or divided government in separation-of-powers systems. Our RDD approach yields compelling evidence that divided government is a cause of delay. We also evaluate and find support for a new hypothesis that divided government is more likely to lead to lead to delay when the personal and political costs that stalemate imposes on politicians are low.

More than 50 years ago, V.O. Key (1964, 688) famously claimed that, "Common partisan control of the executive and legislature does not assure energetic government, but division of party control precludes it." Over time, formal theories of lawmaking have provided support for Key's wisdom, anticipating that divided government will exacerbate the tendency towards stalemate that is already built into America's separation of powers system (cf., Krehbiel 1998; Cameron 2000; Chiou and Rothenberg 2003). Despite these compelling theoretical predictions, the empirical literature disagrees as to how much, and in some instances even whether, divided government matters.

Most of these empirical efforts look for evidence that divided government reduces overall legislative productivity or output. For example, in the seminal book on legislative gridlock, *Divided We Govern* (2005 [1991]), David Mayhew finds that divided government is not associated with a reduction in the number of "landmark bills." Instead, Mayhew notes that shifts in the public's taste for activist government and presidential electoral cycles better account for variation in legislative productivity. Other scholars, however, have uncovered evidence suggesting that fewer new bills are enacted under divided government (Coleman 1999; Cameron 2000; Binder 2003; Chiou and Rothenberg 2003), though these studies tend to disagree about the substantive importance of split partisan control with some even concluding that divided government only matters in the presence of some conditioning variable (cf., Kelly 1993; Howell et al 2000). A handful of scholars have considered legislative productivity at the state level, hoping that the larger number of observations will provide clarity. Again, however, they uncover mixed results, with some efforts finding little to no evidence that divided government lowers productivity (Gray and Lowery 1995; Squire 1998) and others concluding that divided government seems to matter, though again the magnitude of the effect varies notably across studies (Rogers 2005; Bowling and Ferguson 2001; Hicks 2015).

The lack of consensus likely stems from a series of methodological challenges. One of these is accurately measuring legislative productivity—typically, a twofold process that involves identifying significant legislation as well as the public's demand for government action. Disagreements inevitably arise over what constitutes a significant bill, while the demand for legislation is largely unobservable. Though sophisticated techniques have been devised for accomplishing both tasks, they are imperfect and still subject to robust debate (see Binder 2015). Second, the notion that divided government causes gridlock implies a simple model; however, lawmaking is an inherently complex process, particularly in a separation of powers system. In theory, researchers could account for these complexities by including covariates in statistical models of lawmaking, but in practice not all determinants of legislative performance are readily measurable—or known. This raises concerns of both omitted variable and selection biases. Even among those potential determinants of productivity that can be readily measured, there is far from universal agreement as to which should be included in regression models, and issues of sample size (prevalent particularly in studies of the U.S. Congress) limit the the complexity of the statistical tools and the number of explanatory variables that can be employed. Inconsistencies in model specification across studies have further muddied the water when it comes to unearthing a divided government effect.

In light of these challenges, efforts to identify a causal effect of divided government might be best served by shifting focus from overall legislative productivity to more readily operationalized measures of legislative performance. We take this approach and investigate whether split partisan control of government leads to *legislative delay*, that is does it affect the ability of lawmakers to meet critical deadlines? We set our inquiry in the context of state budgeting, considering whether the presence of divided government increases the probability that lawmakers will fail to adopt a new budget prior to the start of a new fiscal year or biennium (see Andersen et al. (2012) and Klarner, Phillips, and Muckler (2012) for other work on the timing of state budget adoption). Though we are hesitant to refer to a late budget as gridlock (since a budget eventually is enacted), such delay does mean that elected officials were unable to make needed policy compromises and decisions prior to a consequential, exogenously-determined deadline that is known to all participants well in advance.

There are several advantages to our focus on fiscal delay. First, budgets are unquestionably significant pieces of legislation. Second, demand in budgeting is clear—a new budget is required by law in every state by the start of each new fiscal year or biennium. Unlike the failure to pass health care or tort reform, for example, the failure to adopt a timely budget is almost universally regarded as an undesirable outcome. The list of potential negative consequences resulting from the failure to adopt an on-time budget include the possibility of a shutdown of state government, reduced services for citizens, and delayed payments to contractors and employees. Third, given the clarity of budget deadlines, it is straightforward to identify delay. We treat delay as occurring whenever a budget is adopted after the start of the new fiscal year.

Our analysis considers budgeting in all states over a 43-year period—1968 through 2010. The large number of observations enables us to implement a regression discontinuity design (RDD), a technique that is increasingly utilized to make causal inferences from observational data. For example, scholars have used RDDs to estimate causal effects for a variety of non-randomly assigned political treatments such as incumbency, partisanship, and ideology (e.g., Lee 2008; Gerber and Hopkins 2011). An RDD allows us to compare the probability of legislative delay in states that are quite similar in terms of their probability of experiencing divided government but differ only in whether divided government is present. This mitigates against the possibility of selection bias, i.e., that some confounder is causing both divided government and delay, while also resolving the thorny need to account for observed or unobserved covariates.

Although RDD is an increasingly common approach to identifying causal effects with observational data, its implementation in separation of powers systems is not straightforward. To make use of an RDD we need to address two challenges, both of which complicate efforts to measure the forcing variable—i.e., the variable which predicts whether an observation will be assigned to treatment (in our case, whether a state will experience divided government). In many political science applications of RDDs, the treatment of interest is the result of a single type of election, meaning that measuring the forcing variable is straightforward. This is not true in our case since the presence or absence of divided government is the result of multiple elections—elections to the governorship, state assembly, and state senate. Our forcing variable must incorporate results for each type of election. A second complication is that a closely divided legislative chamber, in terms of the partisan distribution of seat shares, is not necessarily an indication that the chamber was or is at high risk for a different outcome (e.g., control by the minority party). Partisan gerrymanders and uncontested elections are common in state legislative races and mean that in order to gauge the uncertainty of the partisan control of each chamber, we need to know much more about these elections than just the number of seats each party ultimately won.

To overcome these obstacles, we take cues from the work of Folke (2014) and Fiva, Folke, and Sørensen (2013). Our approach combines the results of thousands of electoral simulations to estimate the probability of divided government. From these estimates we can identify states where the probability of divided government is close to 50 percent. Although we are not the first to employ RDD in such a setting (cf., Bernecker 2014; Feigenbaum, Fouirnaies, and Hall 2016), we are the first to fully account for the electoral processes that conjointly produce divided government. Foreshadowing our results, we find evidence of a causal link between divided government and legislative delay. In states where assignment to divided government is "as-if random," divided government leads to a 12-17% increase in the probability of a late budget. This result is both substantively meaningful and robust to a variety of model specifications.

Our empirical strategy provides the strongest evidence to date of a causal relationship between divided government and legislative performance. Having uncovered this effect it seems natural to ask whether its magnitude differs across contexts. Existing scholarship anticipates that the impact of divided government will be heterogeneous, often varying as a function of the extent of political polarization (e.g., Binder 2003; Krehbiel 1998). We evaluate this widely-held expectation, but also draw upon the state budgeting literature to offer up a new hypothesis. In particular, we argue that divided government is more likely to lead to missed deadlines when the costs (both personal and political) that delay imposes on politicians are low. Because the negative consequences of fiscal impasse vary substantially across states and over time in ways that can be readily operationalized, the context of state budgeting is uniquely situated to evaluate our cost-based theory of legislative performance. Our analyses affirm this hypothesis. When the costs of a missed deadline are high (for example, during election years and in states where a late budget automatically triggers a government shutdown), we find that the frequency of critical delay is low and that there is only modest evidence of a divided government effect. In low-cost contexts, on the other hand, divided government matters a great deal. Indeed, these results suggest that the effect of divided government uncovered by our RDD analysis may be driven largely by the influence of divided government in low-cost circumstances. We also find evidence (albeit weak) that divided government has a greater impact when polarization is high.

Overall, by carefully isolating the effect of divided government and by evaluating a new hypothesis regarding the conditions under which divided government should matter most, we advance ongoing debates about the forces that shape legislative performance. While there may be properties of budgeting that set it apart from other types of legislation (cf. Kousser and Phillips 2012), we believe that these differences do not threaten generalizability to non-fiscal areas. However, even if our findings are limited to budgeting they still stand as an important contribution given the centrality of fiscal policymaking. Importantly, we also make a methodological contribution. By devising a technique that fully accounts for the electoral processes that produce divided government, we facilitate new applications of RDDs to the study of outcomes in separation of powers systems.

1 Divided Government and Legislative Performance

Early work in political science viewed unified partisan control of government as necessary for overcoming the legislative obstacles designed into America's separation of powers system. It is commonly argued that political parties mitigate the challenges of fragmented authority by linking actors across branches. For V.O. Key and many others, the formation of functional linkages, however, is contingent upon the same party occupying a majority of the seats in both legislative chambers and controlling the presidency (Wilson 1911; Key 1964; Pomper 1980). While unified government enables parties to bridge inter-chamber and inter-branch differences, parties are thought to exacerbate these divides during times of split control. Because parties compete, Sundquist (1988) argues, members of one party simply cannot support the proposals of the other. For him, compromise requires partisans to "set aside the principles that are their reason for seeking governmental power," which is likely only under grave conditions "after deadlock has deteriorated into crisis" (629). Work from this tradition anticipates a strong causal linkage between divided government and impasse.

Formal theories of lawmaking also lend support to the expectations of Key, Sundquist, and others, though these theories tend to do so in a more nuanced manner. Krehbiel (1998) and Brady and Volden (1998), for example, develop models of lawmaking that de-emphasize the legislative role of political parties and instead focus on legislative "pivots," the key actors empowered by institutional rules such as the filibuster and the executive veto. In pivotal politics models, status quo policies that lie between the ideological positions of pivotal actors (i.e., within the "gridlock interval") cannot be changed. For scholars working in this tradition, gridlock occurs in the presence of moderate status quo policies and heterogeneous preferences among the pivotal actors, conditions that can arise during both unified and divided government. In practice, however, the gridlock interval should be larger when the legislative and executive branches are controlled by different political parties, thereby making gridlock more common during periods of split partisan control (Cameron 2000).

The empirical literature, however, disagrees as to how much (and even whether) divided government matters. In the first systematic analysis of the link between divided government and legislative productivity, Mayhew (2005 [1991], 4) concludes that "unified as opposed to divided control has not made an important difference in recent times." Relying on both contemporary and retrospective judgments, Mayhew counts important legislation enacted in the post-war period, finding no significant difference in the numbers of landmark laws passed under divided or unified government. In the absence of evidence to support the conventional wisdom, Mayhew considers other factors that shape productivity. While legislators' electoral incentives and norms contribute to constancy, changing public moods, presidential election cycles, or the emergence of cross-cutting cleavages, he argues, can generate variation in productivity.

The null results in Mayhew's work have spawned numerous further inquiries, most of which also focus on legislative productivity. Contrary to Mayhew's original findings, many of these efforts uncover evidence consistent with a divided government effect (Coleman 1999; Howell et al 2000; Cameron 2000; Binder 2003; Chiou and Rothenberg 2003). Across studies, though, there are differences in the extent to which divided government shapes the probability of stalemate. For example, Binder finds evidence that divided government dampens the production of legislation but emphasizes that polarization within each legislative chamber and between-chamber differences in policy preferences matter as much, if not more. Several studies have even concluded that divided government only matters in the presence of some conditioning variable. Kelly (1993), for instance, argues that divided government matters only for the production of innovative policy (which he defines as legislation that is deemed significant by both retrospective and contemporaneous evaluations) while Howell et al. (2000) similarly argue that divided government only decreases the production of landmark legislation (interestingly, though, divided government increases the production of relatively trivial bills).

Other scholars have considered legislative productivity at the state level, hoping to leverage the larger number of observations into greater clarity. Again this work generates inconsistent results, with some efforts finding little to no evidence that divided government lowers productivity (Gray and Lowery 1995; Squire 1998). Among studies that uncover a divided government effect, the effect is typically heterogeneous and often conditioned by the presence of other variables, including the particular configuration of divided government (Rogers 2005), the distribution of seat shares and the amount of cross-party polarization (Hicks 2015), or by policy area (Bowling and Ferguson 2001).

The lack of consensus regarding both the existence and size of a divided government effect likely stems (in part) from the fact that the study of legislative productivity presents significant methodological challenges which researchers approach differently. First, researchers must identify the supply of legislation. There is general agreement that divided government, if it matters, should matter most when it comes to important pieces of legislation. However, there is no clear means for rating the importance of bills, either within a single legislative session or over time. Disagreements exist as to whether one should use the contemporaneous evaluations of journalists, retrospective evaluations of experts, or some more systematic measure of salience (for example, Clinton and Lipinski (2006) suggest the use of a Bayesian item response model).

A second challenge is known as the "denominator problem." As Fiorina (1996) points out, it is difficult to interpret the production of laws in the absence of a measure of demand for policy change. Low levels of output may not necessarily indicate gridlock but rather be a response to low public demand. Unfortunately, since demand is unobservable, measuring it is tricky. Several studies, though, have attempted to do so by incorporating a measure of the policy agenda. Binder (2003) identifies agenda items using *New York Times* editorials and operationalizes gridlock as the number of unresolved issues divided by the total number of agenda items. Others take a different approach, operationalizing legislative productivity as the ratio of significant laws enacted to the sum of important bills that passed and failed (cf., Chiou and Rothenberg 2003; Coleman 1999). These measures are imperfect and subject to concerns of endogeneity. Variation in Binder's measure of demand could reflect changing editorial policies, and all measures of the agenda may be endogenous to legislative characteristics (Chiou and Rothenberg 2008). Even the presence of gridlock itself may indirectly influence agenda size by leaving salient issues unresolved (Binder 2015). Citing concerns over endogeneity, Clinton and Lapinski (2006, 245) note that "it is unclear whether an appropriate denominator is even possible."

Setting aside debates surrounding how to best measure the dependent variable, there is also little agreement as to how researchers should specify a valid model of legislative performance. Given the complexity of the lawmaking process, myriad factors—many known and some probably unknown—are likely to shape legislative productivity. Although researchers diligently try to account for relevant variables, the possibility of unknown determinants of productivity raises the spectre of omitted variable bias. Furthermore, when it comes to constructing empirical models of legislative output, there is a lack of agreement as to which potential determinants ought to be included. This is true even across canonical studies of divided government. For example, Mayhew's (2005 [1991]) model of legislative productivity includes three covariates—indicators for early term presidents (presidents in the first two years) and activist mood (a dummy for the time period of 1961-1976), and a measure of budget surplus or deficit. Binder (2003) draws upon on this work, but adds measures that capture polarization, bicameral differences in preferences, and the size of electoral mandates. She also substitutes an alternative measure of policy mood for the one used by Mayhew. Krehbiel's (1998) analysis departs from the work of both Mayhew and Binder by employing, as its key explanatory variable, change in the size of the gridlock interval. However, Krehbiel also estimates models that include multiple measures of policy mood, as well as movement from divided to unified government and vice versa. Other efforts typically employ some combination of the aforementioned explanatory variables while often adding new ones. Differences in model specification, combined with differences in the operationalization of legislative productivity, make it difficult to interpret both null and and positive results.

Finally, because most studies focus on lawmaking at the national level, the challenges of model specification are exacerbated by issues of sample size. The typical inquiry concentrates on the post-war period, employing congressional sessions as the units of analysis. As a result, empirical tests are run using no more than 25 cases. This limits the complexity of the statistical tools that can be employed as well as the number of explanatory variables that can be included in models, further raising the possibility of omitted confounders. Moreover, the dearth of observations makes it virtually impossible to implement research designs (such as regression discontinuity) that can help unearth causal effects using observational data.

While political science has learned a great deal about divided government from existing research, the methodological challenges, when combined with the mixed findings, suggest that further inquiry is warranted.

2 Reevaluating the Causal Link

We reevaluate the hypothesis that divided government affects legislative performance. In doing so, we utilize a new data source and research design that allow us to avoid existing methodological obstacles and to gain unique causal leverage.

2.1 Late State Budgets as a Measure of Legislative Performance

Most empirical studies focus on lawmaking at the national level, using a measure of legislative output as the dependent variable. We focus instead on legislative delay and examine whether divided government threatens the ability of lawmakers to meet critical deadlines. Specifically, we ask whether the presence of divided government increases the probability that lawmakers will adopt a late budget.

In all 50 states, legislators and governors are required by law to adopt a new budget *prior* to the start of each new fiscal year or biennium. The failure to do so is almost universally seen as an undesirable outcome, making the timeliness of budget adoption a useful measure of legislative performance. In 22 states, a late budget automatically triggers a partial government shutdown, forcing the furlough of public employees, the closing of many state facilities, and reductions in services. Even without a mandated shutdown, a late budget can lead to reductions in service provision, delay payments to public employees, municipalities, and contractors, and compel the state to finance its operations via costly short-term bonds. Late budgets are also unpopular. They typically generate a great deal of negative press coverage, highlighting the costs of delay and the inability of elected officials to "do their jobs." Opinion polls show that late budgets cut into the public approval of both the governor and legislature (Field Poll 2003, 2004; Quinnipiac 2001, 2007; Franklin and Marshall College Poll 2016).

Similar to other types of lawmaking, the adoption of a budget requires intra- and cross-branch negotiations. These negotiations tend to be contested, with key actors trying to ensure that the outcome reflects their priorities. Importantly, the deadline for the adoption of a new budget is set exogenously and does not change from year to year. In other words, it is known well in advance by all participants. A late budget suggests that elected officials were unable to make needed decisions and compromises prior to a meaningful deadline.

From a research design perspective, our focus on budgeting has advantages. Budgets are unquestionably significant pieces of legislation, eliminating the need to rely upon potentially subjective approaches for identifying important bills. Indeed, adopting a new budget is arguably the most essential action in any legislative session (Rosenthal 2004; Kousser and Philips 2012). Unlike in other areas of lawmaking, demand for legislation in the fiscal arena is known—a new budget is required every fiscal year or biennium. In contrast to their counterparts at the national level (who can substitute continuing resolutions for a new budget), state legislators and governors cannot avoid this responsibility. Our measure of delay is also easy to operationalize. For each state and fiscal year, we simply ask, "was the state budget adopted on or before the prescribed deadline?" If the answer to this question is yes, then there is no delay; if the budget is late, we conclude that there is delay.

Moreover by focusing on budgets, we have the added advantage of observing the same type of decisions over time and across states, giving us increased confidence in the comparability of our outcome measures. Existing work, though broader in substantive scope, may suffer somewhat by lumping different types of legislation into a single measure of productivity, especially if the determinants of lawmaking differ across issues and if the types of issues on the legislative agenda vary by year (Howell et al. 2000; Gray and Jenkins 2016).

We should note that there are two recent studies that also investigate the determinants of late budgets (Klarner et al. 2012; Andersen 2012). These efforts consider a wide variety of potential predictors, including the presence of divided government. While both studies find evidence that split control of government matters, they disagree over the extent of its influence. In contrast to our effort, neither unpacks the causal role of any single predictor.

Similarly, there is a small literature on the amount of time it takes Congress to complete key actions. This work uses duration analysis to show that the presence of divided government slows down legislative tasks such as the confirmation of judges, ratification of treaties, and the adoption of important bills (Binder and Maltzmann 2002; Shipan and Shannon 2003; Woon and Anderson 2012; Peake, Krutz, and Hughes 2012; Hughes and Carlson 2015). Like the existing work on late budgets, however, these efforts do not address issues of causality. Furthermore, absent missing a consequential deadline it is unclear whether a count of the amount of time it takes to complete an action is a useful measure of legislative performance.

2.2 Employing A Regression Discontinuity Design

An additional advantage of substantively focusing on state budgeting is that we have a much larger number of observations. This enables us to employ more sophisticated techniques—in this case, a regression discontinuity design (RDD).

As we noted above, one of the many methodological challenges confronting studies of divided government is that the presence of divided government is not randomly assigned, introducing the possibility of endogeneity. States that routinely experience split party control of government may be different from those that tend to exhibit unified party control. While a statistical model can include covariates to account for differences that are both observable and measurable, other unobserved characteristics threaten inference by creating the potential for biased estimation.

An RDD addresses these concerns by allowing us to analyze legislative outcomes in states that are quite similar in terms of their probability of experiencing divided government. Among states where the odds of divided government are close to 50-50, we have a set of observations where assignment to the treatment of interest is effectively as-if random. In theory, these states will be similar in terms of both the observed and unobserved characteristics that influence the likelihood of divided government. By comparing the frequency of legislative delay in states that are quite similar in terms of their probability of experiencing divided government, but that differ in whether divided government is present, we can identify the causal effect of split partisan control on legislative performance. Compared to existing observational research, an RDD greatly reduces the threat of omitted variables, addressing yet another empirical challenge present in the divided government literature.

Indeed, the use of RDDs has become has become common in political science, largely because it allows researchers to draw causal inferences from observational data. Applying this design to the study of divided government, however, faces obstacles pertaining to the creation of our forcing variable—the variable that predicts whether an observation will be assigned to treatment. In most political science applications of RDD, the treatment of interest is the result of a single type of election, meaning that measuring this variable is straightforward. Divided government, however, is the result of multiple election outcomes (elections to the governorship, the state assembly, and the state senate), all of which must be incorporated into our forcing variable.

While it may be tempting to somehow use legislative seat shares as a forcing variable, doing so would give rise to two complications. First, seat shares cannot easily be combined with gubernatorial election results since they each measure distinct quantities of interest. Second, the relationship between seat shares and party control of the state legislature can be ambiguous. Specifically, a chamber that is closely divided (in terms of the partisan distribution of seats) is not necessarily a chamber where the odds of experiencing divided government are close to 50-50. This ambiguity arises because gerrymandered districts and uncontested races are common.[1]

To address these complications, we modify the familiar RDD framework such that results from legislative and gubernatorial elections can be combined into a single forcing variable. To do so we rely on simulations in which electoral shocks of varying magnitudes are administered to real district-level and gubernatorial election results. The simulations give us a sense as to how close a state was, in a given election year, to experiencing a different outcome in terms of divided or unified government. We briefly describe the process of creating our forcing variable here, while providing additional details in the appendix.

The first step in each simulation is to establish the size of a state-level electoral shock (S), the value of which then constrains the size of the district-level shocks (ΔV) that we ultimately apply to real world election results. To get the S, we take a random draw from the uniform distribution (-1 to 1), limiting the magnitude of S to no greater than $\pm 20\%$. The value of S can be either positive or negative, and smaller (larger) values of S produce smaller (larger) values of ΔV.[2] The second step is to take, for each legislative district (j) in state (i), a new random draw (D) from the same uniform distribution as above this time imposing a constraint of $\pm 50\%$. By incorporating (D), we allow for random variation in the size of shocks across districts. Each ΔV_{ij}, then, is a straightforward function of these two draws:

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

[1] In 1999, for example, both Texas and Tennessee had closely divided senates. However, neither state had a single senate race in which the winning margin was less that 10 percentage points, and in both states, nearly one-half of the seats up for election were uncontested.

[2] For example, given Equation 1, a S (i.e., a state shock) of 0.15 will constrain the ΔVs (i.e., district-level shocks) to fall within the range of -0.225 and 0.225, while a S of 0.05 will constrain ΔVs to fall within the range of -0.075 to 0.075.

By design, equation 1 produces district level vote shocks that can never be greater than $\pm 30\%$. We select this range because it encompasses 99% of all of the observed district-level shocks in our elections data (i.e, shocks of greater magnitude are exceedingly rare).[3] To calculate the vote shocks that we apply to gubernatorial elections, we simply treat the entire state as if it were an additional at-large legislative district.[4]

The third step of each simulation is to apply the district level vote shocks. In every district 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 the simulated election. 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.

This process is repeated 40,000 times for every state-election year, noting after each simulation whether the "new" election results produce an outcome that differs in terms of divided or unified government from the real world outcome. After completing all simulations, we then identify, for all state-election years, the smallest state-level vote shock (S) that produces a different outcome in a majority of simulations. We treat this measure as the electoral distance to divided (or unified) government and use it as the forcing variable for our main RDD analyses.[5]

The distribution of our forcing variable is displayed in Figure 2. Note that observations with positive values are assigned to divided government, while observations with negative values

[3]Our RDD results are insensitive to increases in the size of this constraint. However, substantially shrinking the size of the constraint would be problematic. Our approach can only generate a value of the forcing variable when the size of the shock required to produce a different outcome in terms of divided or unified government falls inside the bounds of the constraint. The value of our forcing variable is unaffected if we instead use a normal distribution.

[4]Each simulation incorporates all of the elections in a given state and year. Most states hold gubernatorial and legislative, or "on-term," elections every four years, as well as midterm elections in which voters elect only state legislators. During our time period, about 90% of state general elections fall into one of these categories

[5]There are two alternative approaches one might take to calculating the forcing variable. The first is to utilize the basic approach we develop here, but instead of drawing the state-level shock for S_i from a (constrained) uniform distribution, one could determine S_i using a draw from the distribution of electoral shocks observed over time in state i. This alternative approach, however, should result in a value for the forcing variable that is nearly identical to the one we develop using the uniform distribution. Indeed, our forcing variable is highly correlated ($\rho = 0.995, p < 0.001$) to forcing variable simulated using the distribution of historical electoral shocks within each state. A second alternative is to employ a "centering" procedure (e.g., Wong et al. 2013) that collapses actual vote share margins for the state senate, assembly, and governorship into a single forcing variable. We discuss and apply this approach in the appendix). It is worth noting that one concern we have about centering is that it does not allow for variation in the size of electoral shocks across districts. Again, however, our simulated vote-share forcing variable is highly correlated with a forcing variable generated using centering techniques ($\rho = 0.982, p < 0.001$).

are assigned to unified government. The closer an observation is to the cutpoint (0) the smaller is the state-level vote shock that would be required to produce the opposite outcome. In other words, observations with values close to zero are state election years in which unified and divided government are about equally as likely (and these observations form the core of our RDD analysis). Substantively speaking, a forcing variable of 0.08 means the state experienced divided government but that a state-level vote shift of 8% or greater would have produced unified government. Correspondingly, a value of -0.12 means the state experienced unified government but that a state-level vote shift of 12% or greater would have produced divided government.

The results of our simulations can also be used to create a measure of the probability of divided government. To do so we simply calculate the proportion of simulations for a given state year in which neither the Democratic or Republican party wins control of both legislative chambers and control of the executive branch. Values that are close to zero indicate that divided government is highly unlikely; values close to one indicate that divided government is the expected outcome. The reason we do not rely on this measure as our primary forcing variable is that it is an imperfect predictor of divided government—this is particularly likely for state years in which the probability of divided government is at or near 50% (i.e., those observations that are of primary interest to our investigation).[6]

We do, however, use it as a validity check on our forcing variable. By comparing these two measures, we bolster our assertion that our forcing variable effectively captures the underlying probability of divided government. For example, we want to be sure that where our forcing variable (electoral distance to divided or unified government) is very small, the probability of having divided government is also close to 50-50. Indeed, when we compare these two measures, we observe a strong correlation of 0.965 (for additional details and graphs, please see the technical appendix).

We also conduct an external validity check using data on legislative seat shares. Specifically, we compare values of our forcing variable to the smallest percentage of seats that one party

[6]This issue is analogous to noncompliance in an experimental setting. If this noncompliance is not random, it introduces a threat of bias into our analysis. We could implement a fuzzy RDD, using treatment assignment as an instrument for divided government. However, a fuzzy RDD estimates an even narrower quantity of interest—the effect of treatment on the subset of subjects or observations that comply with treatment assignment.

would need to win in order to move a state from unified to divided control (or vice versa).[7] While (as we noted above) seat margins are an imperfect measure for our purposes, they still tell us something about the vulnerability of the status quo to a different electoral outcome. If our measure actually captures the electoral distance to divided government, we would expect it to be strongly correlated with the analogous seat share measure, and in fact, it is ($\rho = 0.722$). Both of these validity checks give us increased confidence that our measure of distance to divided government is a credible forcing variable.

2.3 Data

Our analysis considers budgeting in nearly all states over a 43-year period—1968 through 2010. The budget data we use are from Klarner et al. (2012). These data identify, for each state fiscal year or biennium, whether the budget was adopted late. In total we have complete budgeting data for 48 states for a total of 1,681 budgets. The exceptions are Alaska, for which we only have 4 observations, and Illinois, for which we only have reliable data on 19 budgets.

Despite the negative consequences of a late budget, fiscal stalemate in the states is fairly common. In our data, approximately 19% of all state budgets were adopted after the required deadline. Unsurprisingly, the frequency of fiscal delay varies across states. A total of 19 states did not experience a single late budget during the time period of our analysis, while over one-quarter of budgets were adopted late in 13 states. The "leaders" in fiscal delay were Wisconsin (90%), New York (86%), and Louisiana (77%).

Figure 1 displays the share of late budgets across all states as well as share of states that experienced divided government. This figure reveals that while there is a great deal of biennium-to-biennium fluctuation in the frequency of late budgets, there is no obvious temporal trend. Indeed, the share of late budgets is just over 19% in the 1970s and 1980s, falling to approximately 18% during the 1990s, and rising to 21% since the year 2000. A cursory glance at the figure does suggest a potential correlation between the presence of divided government and late budgets. The two plotted lines seem to move in tandem in the mid-1980s and from 1999 through the end of our

[7] For example, if Democrats hold 55% of seats in the lower chamber and 60% of seats in the upper chamber, a 5% shift in the lower chamber would result in divided government.

time period. At other points, however, they move in different directions.

The other key data we use are state legislative and gubernatorial election results. For legislative elections we rely upon an ICPSR dataset titled, "State Legislative Election Returns (1967-2010)" (Klarner et al. 2013). These data provide us with the candidate names, party affiliations, and vote counts for state legislative races. The data we use for gubernatorial elections come from the "Voting and Elections Collection" at *Congressional Quarterly*. We supplement these with data on the partisan distribution of legislative seats from Dubin (2007).

2.4 Results of RDD Analysis

We begin by reviewing the identification assumptions of an RDD. The primary assumption is that potential outcomes are smooth across the discontinuity in the forcing variable, which is to say that states just on either side of the threshold between unified and divided government are comparable. In this case, the "no sorting" assumption seems quite plausible. Given that our forcing variable is composed of electoral results for multiple offices, precise control over treatment assignment seems unlikely. However, we do assess the validity of our design in two ways. First, we apply the McCrary (2008) sorting test to assess the density of the forcing variable at the discontinuity. In doing so, we fail to reject the null hypothesis of no sorting. Next, we compare baseline covariates of states that fall near the discontinuity but differ in treatment assignment, and we find no evidence of an imbalance (see the technical appendix for details).

We present our first substantive results in Figure 3, which plots our forcing variable on the x-axis and the observed share of of late budgets on the y-axis. (Keep in mind that observations of our forcing variable with positive values are assigned to divided government, while observations with negative values are assigned to unified government.) Linear regression lines plot the relationship between our forcing variable and outcome of interest. Here, we observe an increase in the probability of a late budget around the cutpoint (0). This suggests that as a state moves from barely having unified to barely having divided government, the probability of a late budget noticeably increases, rising from about 20% to 28%. The figure provides preliminary evidence that divided government has a causal relationship with delay.[8]

To more rigorously examine the effect of divided government we estimate regression models. Ideally, estimation will rely on observations that lie close to the threshold, but in practice, RDD applications commonly rely on alternative global specifications that control for higher-order polynomials of the forcing variable (Imbens and Lemieux 2008). Recent work, however, suggests that this method may produce misleading estimates and advises use of local linear regression (Gelman and Imbens 2014). We follow this advice. Local linear models estimate the local average effect of a treatment at the discontinuity using only those observations that lie within a specified bandwidth on either side of the cutpoint. A challenge with this approach is determining the appropriate bandwidth, a decision that involves a tradeoff between bias and variance. On one hand, wider bandwidths can lead to biased estimates because they incorporate observations further from the discontinuity. On the other, particularly narrow bandwidths can produce unbiased estimates, but a smaller number of observations generally increases the variance of the estimates. Preferences over bandwidth size vary and the choice of bandwidth can be consequential. As a result, some advocate the use of data-driven techniques that minimize researchers' discretion (Imbens and Kalyanaraman 2012; Skovron and Titiunik 2015). Commonly, however, researchers present results for multiple bandwidths.

We present several specifications, the first of which uses the optimal bandwidth computed using the Imbens and Kalyanaraman (I&K) method.[9] This bandwidth is 0.21, which might strike readers as large, especially when compared to existing studies that use vote margins as the forcing variable (in such studies the typical bandwidth employed is 0.05). For this reason we also present regression results using smaller sized bandwidths.[10]

[8]Figure 3 indicates that there is greater variation in the frequency of late budgets among states with unified government. We do not have a good explanation as to why this occurs. It could be that there is something different about states that more often have unified as opposed to divided government or that there are differences in dynamics of budgeting under unified and divided government. Indeed, part of the reason that we opt for an RDD is to minimize the threat of these types of unobserved confounders.

[9]This approach seeks to address the bias-variance tradeoff by minimizing the asymptotic expansion of the mean squared error around the cutpoint.

[10]The algorithm recommended by Calonico, Catteneo, and Titiunuk (2014) produces an optimal bandwidth of 4%. Readers can see results for this bandwidth in Figure 4.

Our first estimations are presented in Table 1 as well as Figure 3. The variable of interest in all models is “Divided government.” Each model in the table is the standard RDD specification and includes our primary variable of interest, the value of the forcing variable, and an interaction between these two. The results shown in Column 1 are generated using the I&K optimal bandwidth. These indicate that moving from barely having unified to barely having divided government leads to an 11.9% increase in the probability of a late budget. This is statistically significant at the 95% level, and substantively quite large. Column 2 uses the more commonly employed bandwidth of 0.05. Here we estimate that divided government leads to a 17.2% increase in the likelihood of a late budget. This effect is also statistically significant, but at the 90% level.

To assess the sensitivity of our results to bandwidth size, we replicate the model shown in Table 1 using bandwidths ranging from 0.03 to 0.20, in increments of 0.01. These results are plotted in Figure 4, where the x-axis is the size of the bandwidth and the y-axis denotes the estimated effect size. The dots represent the point estimates with error bars indicating the 90% and 95% confidence intervals. Estimates of effect of divided government vary somewhat across bandwidths, particularly at bandwidths of less than 0.05. At these small bandwidths our estimates (while always positive and substantively quite large) sometimes fall just short of statistical significance. We suspect that this is due to the relatively small number of observations close to the cutpoint. Overall, however, all specifications indicate that divided government increases the likelihood of delay, and nearly all specifications produce a point estimate that is statistically meaningful at either the 90 or 95% level.[11]

Finally, in the appendix we extend this analysis by considering the effect that divided government has on the length of delay. That is, conditional on a budget being late, is it later during periods of divided government than it is during periods of unified government. Perhaps unsurprisingly, the answer is yes. Our models uncover a divided government effect of between 23 and 28 additional days. Stated somewhat differently, fiscal stalemate lasts, on average, three weeks longer during divided government.

[11]The number of observations that fall within the 3% bandwidth is 185. This grows to 270 at the 5% bandwidth and 520 by the time we get to a 10% bandwidth.

2.5 Alternative Estimation Strategies

We believe that our RDD results provide strong evidence of a causal link between divided government and delay. As we have previously noted, the strength of RDD (given our emphasis on causal identification) is that it mitigates the threat of endogeneity, particularly concerns of selection bias and omitted variable bias. Nonetheless, there are other options that have the potential to address these issues, for example inverse probability weighting (IPW) and fixed effects (FE) models. However, we believe that RDD is the most suitable approach to identifying a causal effect of divided government because its identification assumptions are transparent and more readily testable. As previously noted, we have used several formal tests to support the validity of our RDD (see Section 2.4 and appendix for details). That being said, we still estimate models here using both the IPW and FE approaches. Before considering the results, though, it is important to note that each of these alternatives estimate a different quantity of interest than the RDD. While the RDD produces a local average treatment effect, IPW models make use of all observations to estimate an overall average treatment effect, and the FE models estimate the average within-state effect of divided government. These differences complicate comparisons across models.

First, we utilize inverse probability weighting (IPW). Although less common as a method for causal inference in political science, researchers in statistics and biostatistics have long employed IPW to estimate treatment effects using observational data (Curtis et al. 2007; Gylnn and Quinn 2010; Blackwell 2013). This approach weights each observation by the inverse of its probability of assignment to its observed treatment condition.[12] An IPW design rests on the relatively strong assumptions that the probability of treatment is properly specified and that the treatment effect is constant. The statistical models for estimating propensity scores are subject to familiar concerns about misspecified models and omitted variables. Failure to account for potential confounders could bias results and undermine causal claims. Rather than using propensity scores, we

[12] Observations that experience divided government are weighted by the inverse of the probability of divided government, while observations that experience unified government are weighted by the inverse of the probability of unified government (1 - probability of divided government).

make use of the estimated probabilities of divided government generated by our simulations.[13] By doing so, we hope to avoid concerns over model specification and omitted variables, though there is no way to be certain that we do so.

Table 2 reports our IPW results. For the sake of comparability, we present a regression model (Column 1) analogous to our RDD analysis—a bivariate OLS model that only includes divided government. These results confirm a divided government effect: the coefficient of interest is positive, substantively meaningful, and statistically significant (at the 99% level), indicating that split partisan control of government increases the probability of a late budget by approximately 9 percentage points. This point estimate remains nearly the same if we estimate covariate adjusted models.

Second, we consider FE models. These models are especially useful for accounting for unobserved time-invariant characteristics of states. For example, late budgets are more common in some states than others, which may raise concerns that unmeasured factors, such as institutions or norms, may shape the timing of budget adoption. However, they are not impervious to threats of endogeneity. FE models, for example, cannot account for unobserved variables that change over time or the possibility that the implications of fixed state characteristics themselves change over time. Another limitation of FE models is that by estimating only within-state effects, these specifications discard information about differences between states.

We use fixed effects in two ways. We begin with Table 2 which reports two new specifications of our IPW model, one that includes state fixed effects (Column 2) and another that includes both state and year fixed effects (Column 3). The addition of fixed effects increases the overall explanatory power of our IPW model, but does not change its substantive conclusions. The coefficient on divided government is only slightly smaller (approximately 8 percentage points instead of 9) and remains statistically significant at the 99% level.

Table 3 reports the results of basic FE models (i.e., models that do not use weights or

[13]This means that in our application the weights are the inverse of the estimated probability of divided (unified) government. Thus, states that are less (more) likely to experience divided government are weighted more (less) heavily. Intuitively, a state that has a low probability of assignment to treatment but actually experienced divided government will serve as a proxy for other states under unified party control.

covariates to account for the probability of divided government). Column 1 reports the results of a specification that uses only state fixed effects, while Column 2 reports estimates from a model utilizing both state and year fixed effects. These results are broadly consistent with our main findings that divided government increases the likelihood of legislative delay. These basic FE models, though, produce substantially smaller estimates of the impact of divided government, about 1/2 the size reported in our IPW estimations. Here the coefficients on divided government are only statistically significant at the 90% level.

For reasons we note above, we need to be careful when comparing results across models. However, it is clear that we uncover a larger divided government effect when we directly account for the probability of divided government (as we do in both the RDD and IPW approaches). This suggests to us that fixed effects alone may not be sufficient to account for selection effects. Finally, although neither the fixed effect nor IPW models can give us the same causal leverage as the RDD analysis, the fact that we observe a similar pattern across all models increases our confidence that divided government is indeed a cause of legislative delay.

3 Looking for Heterogeneous Effects

Having uncovered strong evidence of a causal relationship between divided government and delay, it is natural to ask whether this effect is heterogeneous, that is, whether there are circumstances that either magnify or mitigate against the effect of split partisan control of government. Existing work often anticipates that divided government's impact will be shaped by the degree of cross-party political polarization (e.g., Binder 2003; Krehbiel 1998). We evaluate this expectation but also offer a new hypothesis. In particular, we argue that the effects of divided government should depend, in part, upon the costs that delay imposes on politicians. We reason that when the costs are high, elected officials have a greater incentive to reach legislative bargains even if doing so requires compromise. These costs should mitigate the tendency toward stalemate inherent in divided government.[14]

[14] We note that several studies of gridlock argue that the effect of divided government varies as a function of the characteristics of the legislation under consideration (e.g., Howell et al 2000). By focusing exclusively on budget adoption, our data do not include legislation of varying significance and therefore are not well suited to evaluate this

Delay can impose two types of costs on lawmakers—political and private. In state budgeting, the political costs of delay come in the form of public and interest group anger over a late budget. As we noted above, late budgets generate negative press coverage and result in lower public approval of elected officials, potentially jeopardizing their chances at reelection. In contrast, private costs come from the fact that the legislature must stay in session until a final agreement with the governor is reached, crowding out other activities that lawmakers value (Andersen et al 2102; Klarner et al 2012).

The price of fiscal delay to politicians, however, should differ across states and over time and in ways that are readily operationalized, making the budgetary arena an ideal setting for evaluating our cost-based hypothesis. The political costs of delay should be highest in those states that require a partial shutdown of government if the budget is late. In states with a shutdown rule, delay results in the closing of many state facilities and parks, the furlough of public employees, and the suspension of "nonessential" services (Pulsipher 2004). In jurisdictions without such a rule, lawmakers can finance government operations through either continuing resolutions or some combination of reserve funds, IOUs, intergovernmental revenues, borrowing, and deferrals of expenditures. While these approaches are short-term, they help insulate elected officials from the anger of voters and interest groups.[15] The political costs of delay should also be shaped by the proximity to the next election—a late budget during an off year is probably less likely to be remembered on election day than one that has more recently occurred.

Private costs, in contrast, should vary as a function of legislative institutions. Because legislative service in most states is not a full-time (or well-paying) job, many lawmakers maintain careers outside of government. Protracted budget negotiations, which often require lawmakers to stay in the capitol long after the regular legislative session is scheduled to end, can be personally costly to those legislators who must return home for professional reasons. This should be especially true for lawmakers in states that traditionally meet in short legislative sessions, i.e., sessions that

hypothesis.

[15]The presence of a shutdown rule is almost always determined by state constitutional law and is not easily manipulated.

are scheduled to end prior to the start of the new fiscal year.

To evaluate our cost-based hypothesis, we begin by replicating our RDD analysis on subsets of data for states where the costs of delay are high or low, focusing on the three factors mentioned above. While our ability to make causal claims about heterogeneous effects is limited, we can nonetheless provide suggestive evidence as to how divided government may operate in varied contexts.

The results are presented in Figure 5. Each panel displays the point estimates and confidence intervals from our RDD model for three different bandwidths—0.05, 0.10, 0.20. The first column (labeled "Higher Cost") shows the effect of divided government under conditions that should increase the cost of stalemate. The second column (labeled "Lower Cost") shows the effect of divided government under lower cost conditions, and the third column shows the difference. Generally, we find a consistent pattern: split partisan control of government is most consequential when the costs of delay are low. Though not all of these differences reach conventional levels of statistical significance, collectively, they provide compelling evidence of a heterogeneous effect.

We begin with the top panel of Figure 5 which shows the impact of divided government in the presence and absence of a shutdown rule. In states with such a requirement, the effect is quite small (at all bandwidths) and fails to approach conventional levels of statistical significance. In states without such a requirement, however, it is large and consistently significant at either the 95% or 90% levels. At the 5% bandwidth, for example, moving from just barely having unified to just barely having divided government results in a whopping 35% increase in the probability of a late budget (compared to a 2% increase in states with a shutdown rule).

The second panel considers states with and without long legislative sessions. Here we define higher cost states as those with sessions that end prior to the start of the new fiscal year. Correspondingly, lower cost states have legislative sessions that extend beyond the budget deadline. Again, we find the effect of divided government is much greater in lower cost states. In states with short sessions, divided government has only a small effect (ranging from 0 to 6%) that is statistically indistinguishable from zero. In contrast, divided government is both substantively and statistically significant in states with longer session lengths, ranging from 21% to 30%.

The bottom panel considers the consequences of split partisan control on stalemate during election and non-election years. Consistent with our results above, divided government has a larger impact during non-election years. When states budget during an election year, moving from unified to divided government has a small and statistically insignificant effect on the probability of adopting a late budget. During non-election years, the size of this effect is statistically meaningful and substantively large, ranging from a low of 13% to a high of 22%.

Next, we examine the role of political polarization, though our efforts are constrained somewhat by issues of data availability. In particular, the best existing measure of state-level polarization is that of Shor and McCarty (2011), which is only available for 1993-2010. We utilize this measure, but doing so considerably reduces our number of observations. For these reasons, we also employ an alternative approach in which we simply split our time period in half, estimating the effect of divided government from 1968 through 1989 and then from 1990 through 2010. Existing research shows that polarization during the first half of our time series was much lower than polarization in the second half (McCarty, Poole, and Rosenthal 2006). We believe that an increase in the effect of divided government on stalemate in this latter period would at least be suggestive evidence that polarization exacerbates the effects of divided government.

The top panel of Figure 6 compares the effect of divided government in states in the bottom and top quartiles of political polarization, using the Shor and McCarty measure. The results are consistent with expectations, though they are not particularly compelling (especially at smaller bandwidths). Using a 20% bandwidth, the effect of divided government is substantively much larger (35 percentage points) in high-polarization states. This difference falls just short of statistical significance ($p = 0.12$ using a two-tailed test). At narrower bandwidths the difference in the effect of divided government shrinks substantially, even reversing itself at the 5% bandwidth. However, these small-bandwidth models are estimated with fewer than 100 observations.

The bottom panel of Figure 6 compares the effect of divided government during an era of lower polarization (1968 through 1989) to its effect during an era of relatively high polarization

(1990 through 2010). In the post-1990 period, the effect of divided government is substantively and statistically significant across all bandwidths, ranging from a low of 15% to a high of 23%. In the earlier low-polarization era, the estimated effect of divided government is much smaller, ranging from just 2% to 14%. Neither the estimates for the lower polarization period nor the differences between time spans, however, reach statistical significance.[16]

As with our main analysis, we also estimate heterogeneous effects using IPW and FE models. These results are analogous to those presented above, providing further evidence that the effect of divided government is heterogeneous, especially with respect to the costs of delay. To conserve space, we present these results in the appendix.

4 Discussion

Empirical work in political science has raised questions about the presumed link between the presence of divided government and legislative performance. We revisit this debate, with a focus on establishing a causal relationship. Rather than relying on difficult-to-operationalize measures of legislative productivity, we instead consider whether divided government affects the ability of lawmakers to meet critical deadlines. Specifically, whether divided government makes it more likely that state legislators and governors will fail to adopt a budget prior to the start of the new fiscal year (as is required by state law). By focusing on delay instead of productivity we avoid measurement problems, particularly the challenges inherent in measuring the supply of and demand for legislation. Our substantive focus also provides a large number of observations, which we leverage to design and implement a simulations-based RDD approach that is suitable for separation of powers systems. Using this approach we find that divided government does indeed affect the likelihood of critical delays. These effects are statistically significant and substantively quite

[16]We also considered whether the effect of divided government differs in states with and without the line item veto and whether the effect varies as a function of the particular type of divided government experienced (split branch vs. split legislature). When we employ small bandwidths we find some very modest evidence that late budgets are less likely when governors *do* posses item veto power. This result, however, does not approach statistical significance. We should note that since governors in 88% of states possess the item veto power, there are very few observations with which to estimate the effects of divided government in non-item veto states (especially when we employ small bandwidths). With respect to split-branch vs. split legislature forms of divided government we do not find any evidence that the probability of delay differs by type of divided government.

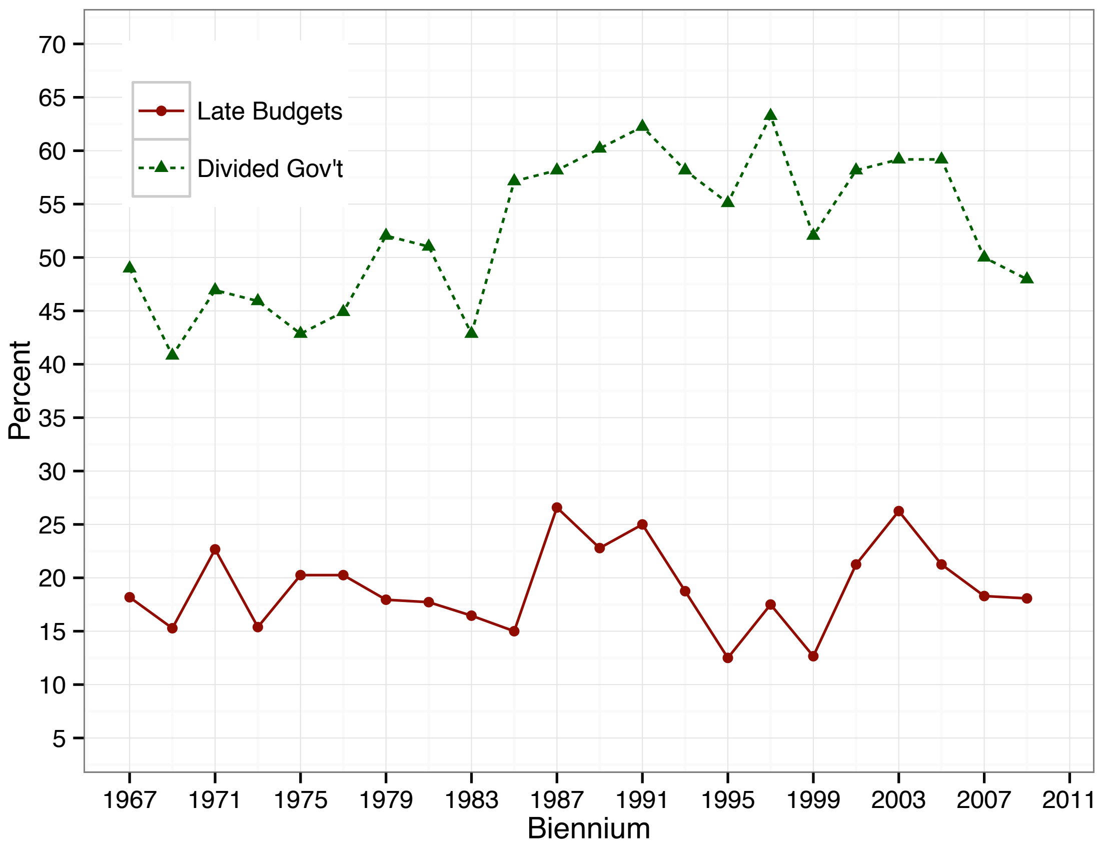
meaningful. Our RDD analyses provide the strongest evidence to date that there exists a strong causal relationship between the presence of divided government and legislative performance.

We also evaluate a new hypothesis that the consequences of split partisan control of government vary as a function of the costs that delay imposes on lawmakers. Our analyses indicates that divided control matters most when the private and political costs of stalemate are low—states where the legislature routinely meets in long regular sessions, states that do not require a partial shutdown of government in the presence of a late budget, and in years without legislative or gubernatorial elections. In bargaining contexts where the costs of impasse are high, however, the effect of divided government is negligible. Our results suggest that the average effect of divided government uncovered by our RDD analyses may be driven largely by the influence of divided government in low-cost circumstances.

While we are not alone in arguing that the consequences of divided government may be context specific, we believe that we are the first to demonstrate that the costs of impasse play an important role in shaping its effects. Uncovering this result is clearly aided by our decision to focus on fiscal policy making, an arena in which the costs of impasse can be readily operationalized. Our finding is useful in predicting what types of legislation may be be most and least likely to be impacted by the presence of divided government. Notably, the results we have presented here also raise the possibility that the failure to properly account for heterogeneous effects may explain the null results in existing studies of divided government and legislative performance.

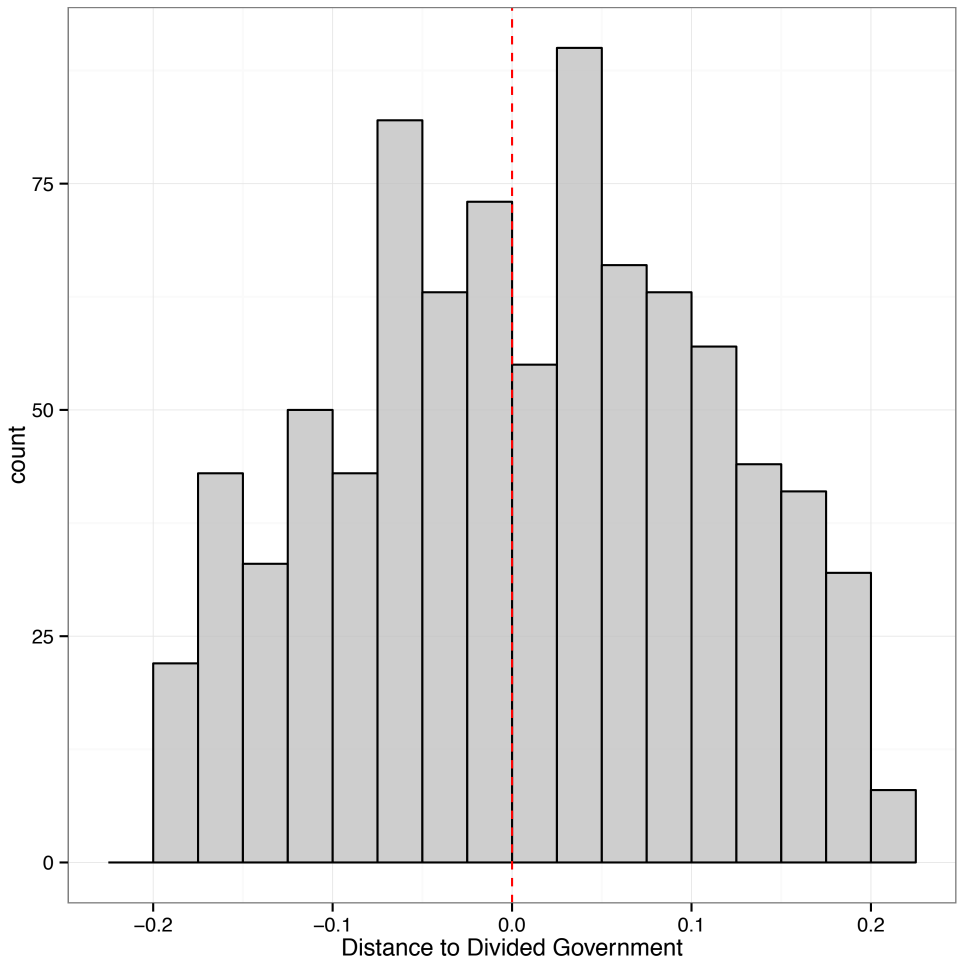
Similarly, our results may help account for the finding of Mayhew and others that divided government does not necessarily lead to a reduction in the passage of landmark legislation. If the issues addressed by landmark legislation are perceived by voters or key interest groups as being particularly important, the costs of inaction may be enough to motivate lawmakers of different parties to compromise. Finally, our analyses suggest that there may be institutional steps that can be taken to decrease the frequency of stalemate (besides, of course, abandoning a separation of powers system entirely). The political costs of inaction could be increased, for example, by more frequently using sunset provisions, exploding deadlines, or government shutdown requirements in the legislative process. Alternatively, and perhaps more appealing, are steps that could increase the private costs of stalemate to lawmakers. These costs could arguably be enhanced by fining or docking the pay of lawmakers or mandating attendance when certain types of legislation are delayed.

Figure 1: Frequency of Late State Budgets & Divided Government, 1968-2010



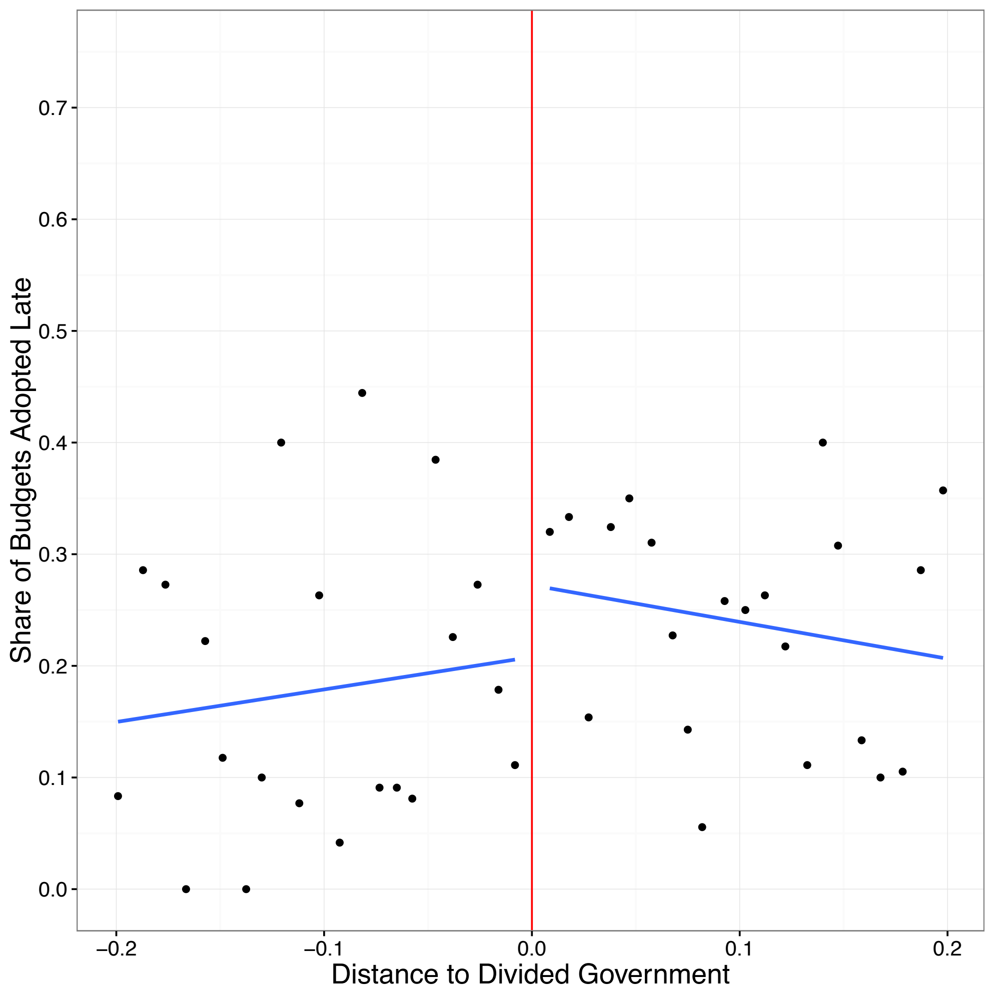
Note: This graph depicts the proportion of state governments that are divided as well as the share of budgets that are adopted late per biennium. Each biennium begins with an odd-numbered year.

Figure 2: Distribution of the Forcing Variable



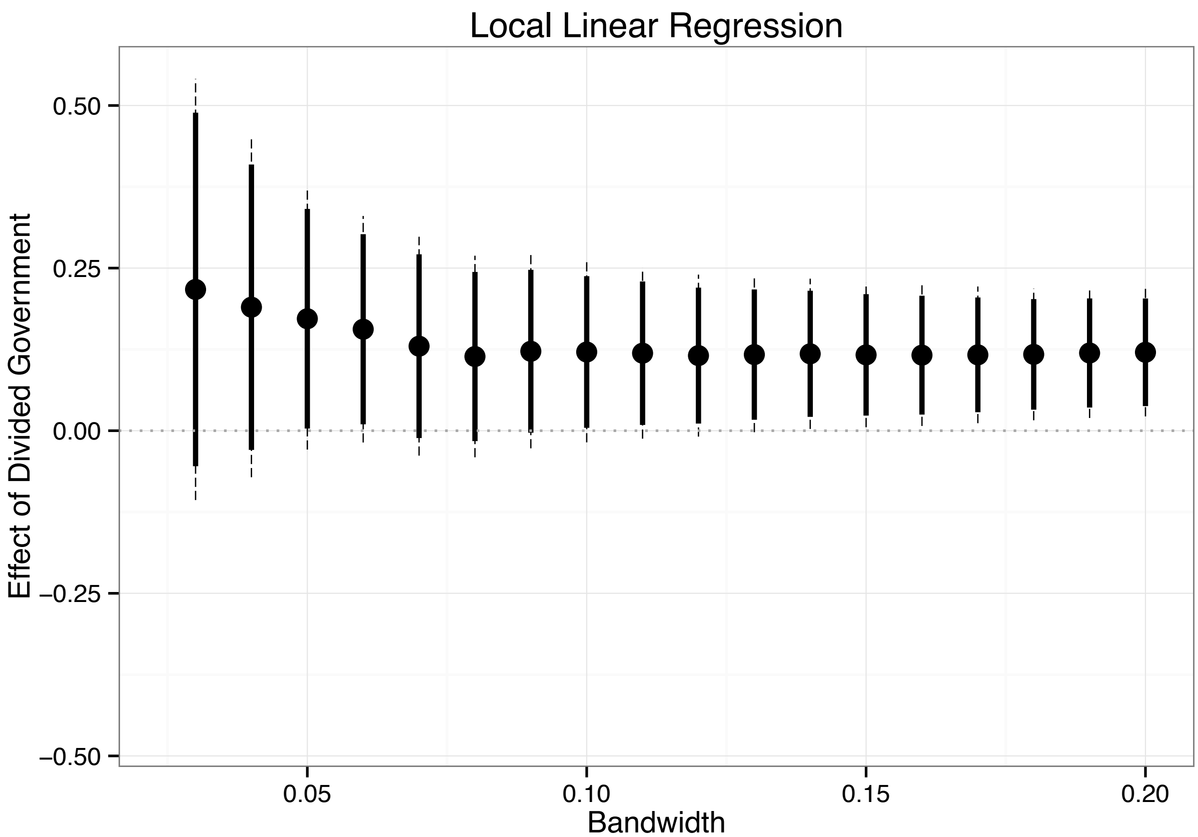
Note: The histogram displays the distribution of the forcing variable using bins of 5%. Zero on the x-axis is the cutpoint. Observations to the right of the cutpoint (i.e., positive values) have divided government; observations to the left of the cutpoint (i.e., negative values) have unified government.The y-axis is a count of the number of state years that fall into each bin.

Figure 3: Late Budgets & Divided Government



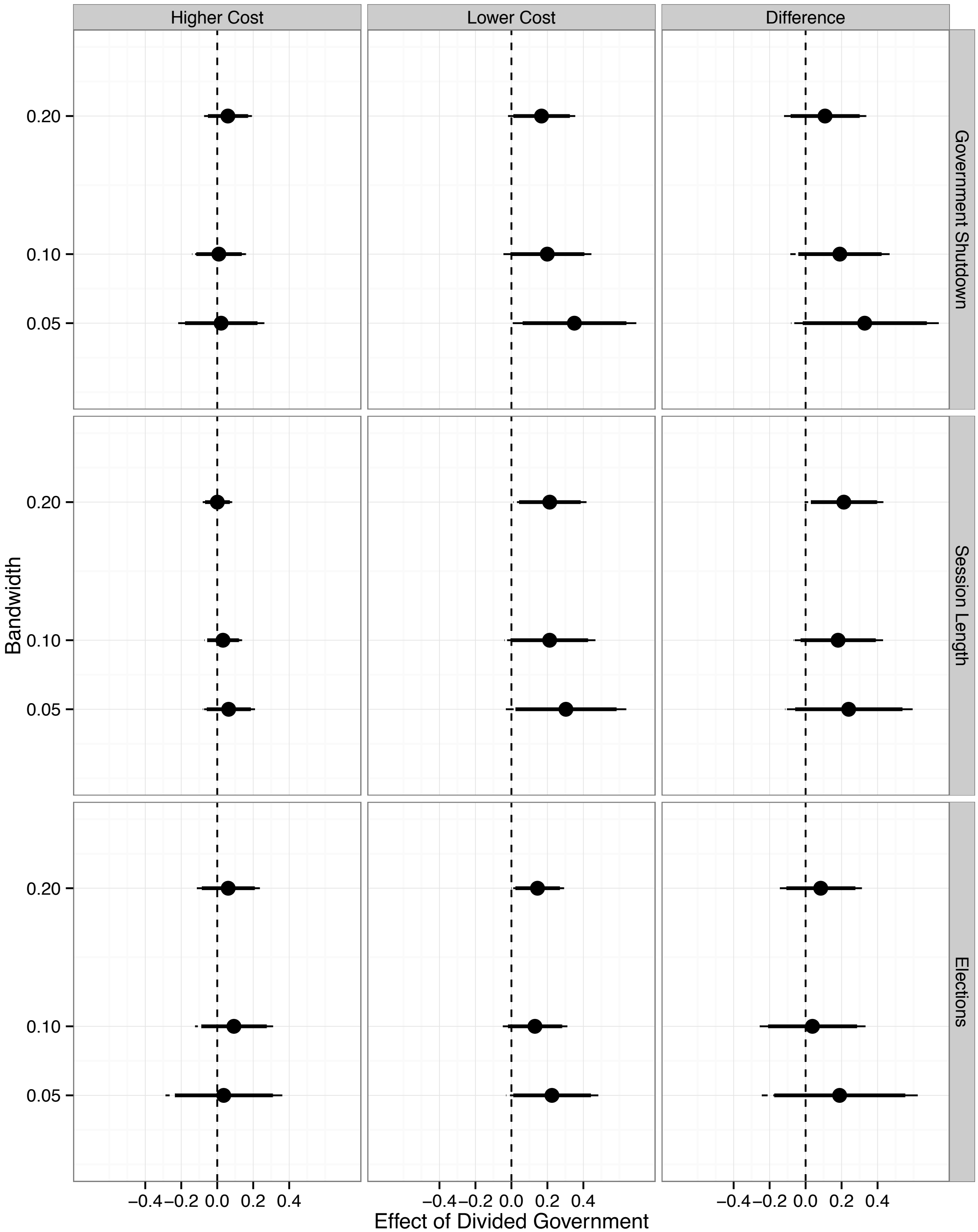
Note: This graph plots the share of late budgets against the forcing variable. The x-axis is the distance to divided government centered at 0, and the y-axis is the share of budgets adopted after the start of the fiscal year. The points are averages of the number of late budgets in 1% bins.

Figure 4: The Effect of Divided Government



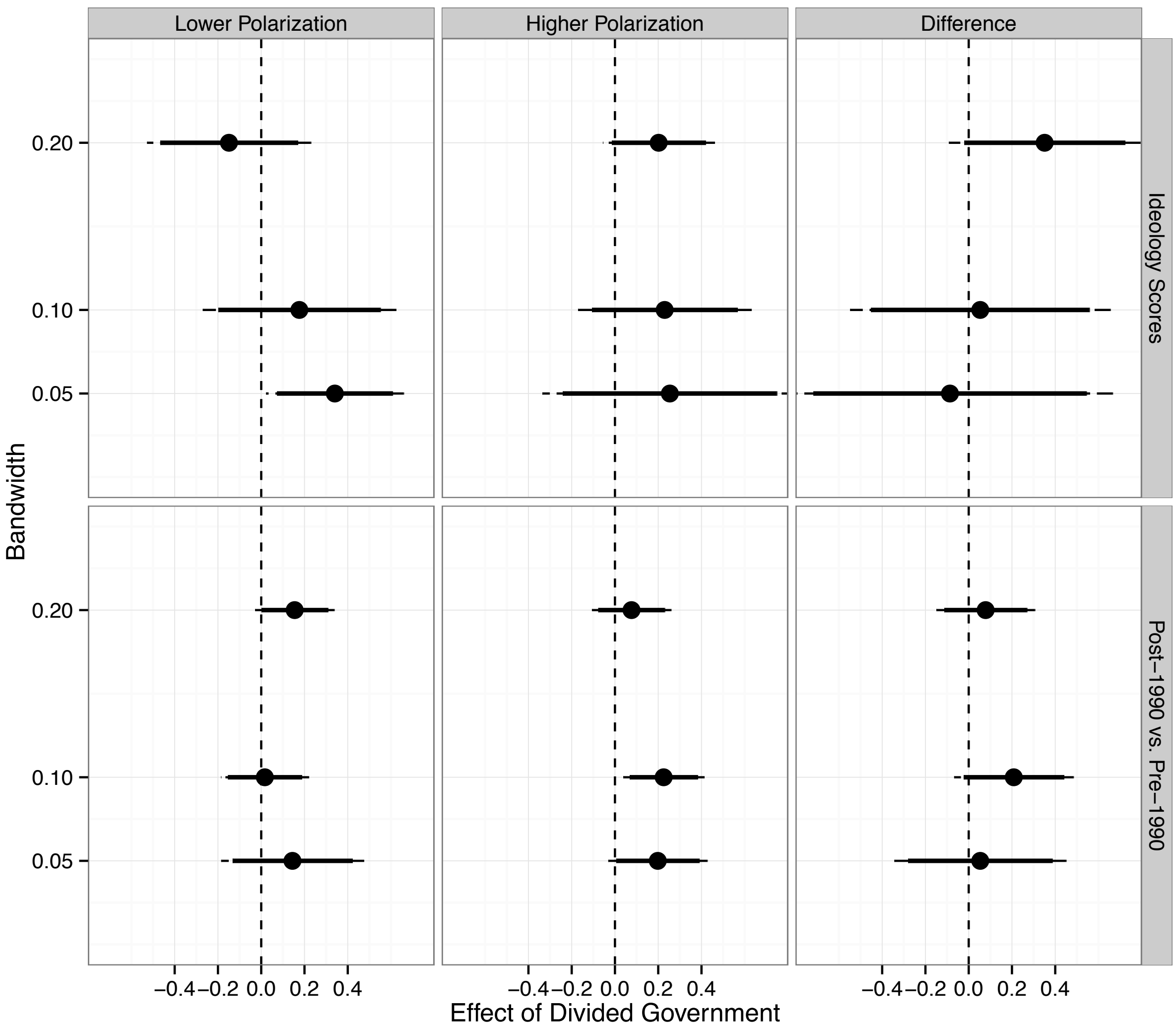
The figure plots the effect of divided government across multiple bandwidths. The horizontal axis measures the bandwidth size, and the vertical axis measures the effect size. The dots indicate point estimates from local linear regression models, and the error bars reflect two-tailed tests. The solid black lines show 90% confidence intervals while the dashed lines indicate 95% confidence intervals.

Figure 5: Heterogeneous Effects of Divided Government (RDD Results)



This figure summarizes the heterogeneous effects of divided government. The x-axis measures the effect size while the y-axis indicates the bandwidth. The black dots indicate point estimates from local linear regression models. Error bars illustrate 90% and 95% confidence intervals using 2-tailed tests.

Figure 6: Heterogeneous Effects of Divided Government (RDD Results)



This figure summarizes the heterogeneous effects of divided government. The x-axis measures the effect size while the y-axis indicates the bandwidth. The black dots indicate point estimates from local linear regression models. Error bars illustrate 90% and 95% confidence intervals using 2-tailed tests.

Table 1: The Effect of Divided Government: RDD Results

	Dependent Variable: Late Budget	
	Bandwidth	
	0.21	0.05
	(1)	(2)
Divided government	0.119	0.172
	(0.058)	(0.103)
Distance to divided gov't	0.236	−5.110
	(0.413)	(3.319)
Distance to divided gov't * Divided government	−1.046	6.421
	(0.608)	(4.612)
Constant	0.194	0.091
	(0.039)	(0.071)
Observations within Bandwidth	859	270
R^2	0.014	0.025
Adjusted R^2	0.010	0.014
Residual Std. Error	0.319 (df = 855)	0.288 (df = 266)
F Statistic	4.017 (df = 3; 855)	2.243* (df = 3; 266)

Note:
Maximum of robust and conventional standard errors reported. Two-tailed test.

Table 2: The Effect of Divided Government: IPW Results

	Dependent variable:		
	Late Budget		
	(1)	(2)	(3)
Divided government	0.093	0.081	0.081
	(0.030)	(0.028)	(0.030)
Constant	0.176	−0.033	−0.014
	(0.018)	(0.024)	(0.076)
Fixed Effects	No	State	State and Year
Observations	1,006	1,006	1,006
R^2	0.012	0.475	0.511
Adjusted R^2	0.011	0.450	0.465
Residual Std. Error	0.505 (df = 1004)	0.377 (df = 959)	0.372 (df = 918)
F Statistic	12.574 (df = 1; 1004)	18.879 (df = 46; 959)	11.041(df = 87; 918)
Note:	OLS results with robust standard errors. Two-tailed test.		

Table 3: The Effect of Divided Government: Fixed Effects Results

	Dependent variable:	
	Late Budget	
	(1)	(2)
Divided government	0.039	0.036
	(0.020)	(0.021)
Constant	0.107	0.058
	(0.009)	(0.087)
Fixed Effects	State	State and Year
R^2	0.455	0.470
Adjusted R^2	0.438	0.439
Residual Std. Error	0.297 (df = 1595)	0.297 (df = 1553)
F Statistic	27.160 (df = 49; 1595)	15.159 (df = 91; 1553)
Note:	OLS results with standard errors clustered by state. Two-tailed test	

References

Andersen, Asger Lau, David Dreyer Lassen, and Lasse Holbøll Westh Nielsen. 2012. "Late Budgets." *American Economic Journal: Economic Policy* 4(4): 1-40.

Bernecker, Andreas. 2014. "Divided we Reform? Evidence from U.S. Welfare Policies." Manuscript, University of Mannheim.

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.

Blackwell, Matthew. 2013. "A Framework for Dynamic Causal Inference in Political Science." *American Journal of Political Science* 57(2): 504-520.

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.

Brady, David W. and Craig Volden. 1998. *Revolving Gridlock: Politics and Policy from Jimmy Carter to George W. Bush*. Boulder, CO: Westview.

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

Cameron, Charles M. 2000. *Veto Bargaining: Presidents and the Politics of Negative Power*. Cambridge: Cambridge University Press.

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

Chiou, Fang-Yi and Lawrence S. Rothenberg. 2007. “Comparing Legislators and Legislatures: The Dynamics of Legislative Gridlock Reconsidered.” *Political Analysis* 16(2): 197-212.

Clinton, Joshua D. and John S. Lapinski. 2006. “Measuring Legislative Accomplishment, 1877-1994.” *American Journal of Political Science* 50(1): 232-49.

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

Curtis, Lesley H., Bradley G. Hammill, Eric L. Eisenstein, Judith M. Kramer, and Kevin J. Anstrom. 2007. “Using Inverse Probability-Weighted Estimators in Comparative Effectiveness Analyses with Observational Databases.” *Medical Care* 45(10): S103-S107.

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

Feigenbaum, James J., Alexander Fouirnaies, and Andrew B. Hall. 2016. ”The Majority-Party Disadvantage: Revising Theories of Legislative Organization.” Forthcoming, *Quarterly Journal of Political Science*.

Field Poll. 2003. *Release #2081*. San Francisco, CA: Field Research Corporation.

Field Poll. 2004. *Release #2143*. San Francisco, CA: Field Research Corporation.

Fiorina, Morris. 1996. *Divided Government*. 2d ed. Boston: Allyn Bacon.

Fiva, Jon H., Olle Folke, and Rune J. Sørensen. 2013. “The Power of Parties.” CESifo Working Papaer Series No. 4119

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

Franklin & Marshall College. 2016. "January 2016 Franklin & Marshall College Poll." Lancaster, PA: Center for Opinion Research, Franklin & Marshall College.

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.

Glynn, Adam N., and Kevin M. Quinn. 2010. "An Introduction to the Augmented Inverse Propensity Weighted Estimator." *Political Analysis*. 18(1): 36-56.

Gray, Thomas R. and Jeffrey A. Jenkins. 2016. "Pivotal Politics and the Ideological Content of Landmark Laws." Unpublished manuscript, University of Virginia.

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.

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

Howell, William, Scott Adler, Charles Cameron and Charles Riemann. 2000. "Divided Government and the Legislative Productivity of Congress, 1945-94." *Legislative Studies Quarterly* 25(2): 285-312.

Hughes, Tyler and Deven Carlson. 2015. "Divided Government and Delay in the Legislative Process: Evidence from Important Bills, 1949-2010." *American Politics Research* 43(5): 771-92.

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.

Kelly, Sean. Q. (1993). “Divided We Govern? A Reassessment. *Polity* 25(3): 475-484.

Key VO. 1964. *Politics, Parties and Pressure Groups*. 5th edition. New York: Ty Crowell

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

Klarner, Carl. E., Justin H. Phillips, and Matt Muckler. 2012. “Overcoming Fiscal Gridlock: Institutions and Budget Bargaining.” *The Journal of Politics* 74(04): 992-1009.

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

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

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

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

Mayhew, David R. 2011. *Partisan Balance: Why Political Parties Don’t Kill the US Constitutional System*. Princeton, NJ: Princeton University Press.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge, MA: MIT Press.

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

Peake, Jeffrey S., Glen S. Krutz, and Tyler Hughes. 2012. “President Obama, the Senate, and the Polarized Politics of Treaty Making.” *Social Science Quarterly* 93(5): 1295-1315.

Pomper, Gerald. 1980. “The Contribution of Poltical Parties to American Democracy” in G. Pomper (Ed.), *Party Renewal in America: Theory and Practice*. New York: Praeger.

Quinnipiac University. 2001. “Senator Clinton’s Approval Tops 50% for First Time, Quinnipiac University Poll Finds.” Hamden, CT: Quinnipiac University.

Quinnipiac University. 2007. “More Pennsylvania Voters Blame Lawmakers for Gridlock, Quinnipiac University Poll Finds.” Hamden, CT: Quinnipiac University.

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.

Shipan, Charles R. and Megan L. Shannon. 2003. “Delaying Justice(s): A Durationa Analysis of Supreme Court Confirmations.” *American Journal of Political Science* 47(4): 654-68.

Shor, Boris, and Nolan McCarty. 2011. “The Ideological Mapping of American Legislatures.” *American Political Science Review* 105(03): 530-551.

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.

Sundquist, James L. 1988. ”Needed: A Political Theory for the New Era of Coalition Government in the United States.” *Political Science Quarterly* 103(4): 613-635.

Wilson, Woodrow. 1911. *Constitutional Government in the United States*. New York: Columbia University Press.

Wong, Vivian C. Peter M. Steiner, and Thomas D. Cook. 2013. “Analyzing Regression-discontinuity Designs with Multiple Assignment Variables: A Comparative Study of Four Estimation Methods.” *Journal of Educational and Behavioral Statistics* 38(2): 107-141.

Woon, Jonathan and Sarah Andersen. 2012. “Political Bargaining and the Timing of Congressional Appropriations.” *Legislative Studies Quarterly* 37(4): 409-36.

Appendix: Is Divided Government A Cause of Legislative Delay?

Patricia A. Kirkland
Center for the Study of Democratic Institutions
Vanderbilt University
patricia.a.kirkland@vanderbilt.edu

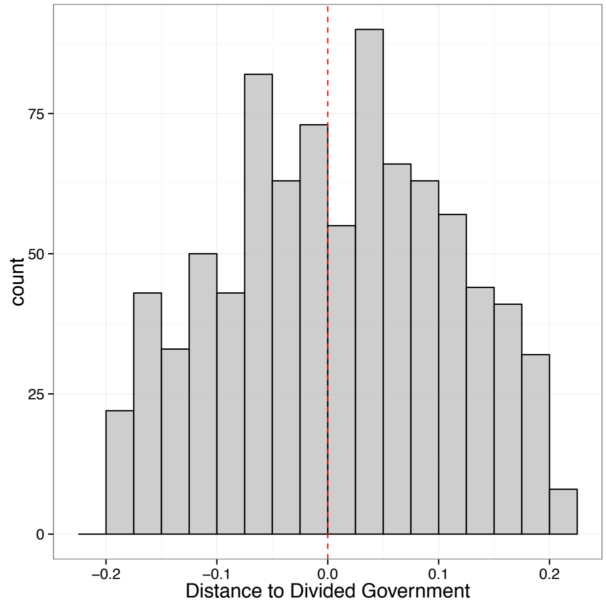
Justin H. Phillips
Department of Political Science
Columbia University
jhp2121@columbia.edu

August 21, 2017

A Validity of the RDD

The "no sorting" assumption—that potential outcomes are smooth across the discontinuity—is the key identifying assumption of the RDD. We test the validity of our RDD in several ways. First, we examine the distribution of the forcing variable as shown in Figure A1. Then, we use the McCrary (2008) sorting test to formally assess the density of the forcing variable at the threshold. In doing so, we fail to reject the null hypothesis of no sorting (log difference in heights is 0.037 with SE 0.242; p = 0.878). We also conduct a series of placebo tests, using pre-treatment covariates as dependent variables to check for discontinuities at the threshold of the forcing variable. As in our RDD analysis, we use local linear regression models and rely on the I & K optimal bandwidth. The results are displayed in Table A1 and provide further support for the validity of the RDD. Covariates include indicators of government shutdown provisions, biennial budgeting, and supermajority budget passage requirements, as well as measures of legislative session length and state-level economic conditions.

Figure A1: Distribution of the Forcing Variable



Note: The histogram displays the distribution of the forcing variable using bins of 5%. 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.

B Our Estimates of the Probability of Divided Government

Recall that we use our simulations to construct two measures. The first is our rating variable—distance to divided (unified) government—which is the smallest state-level vote shock that produces a different outcome in terms of divided or unified government in a majority of simulations. The second measure that we construct is the probability of divided government. The probability measure is simply the proportion of simulations that produce divided government in a given state year. Figure A2 plots these two measures against one another. As the figure suggests, these two measures are highly correlated ($\rho = 0.964$).

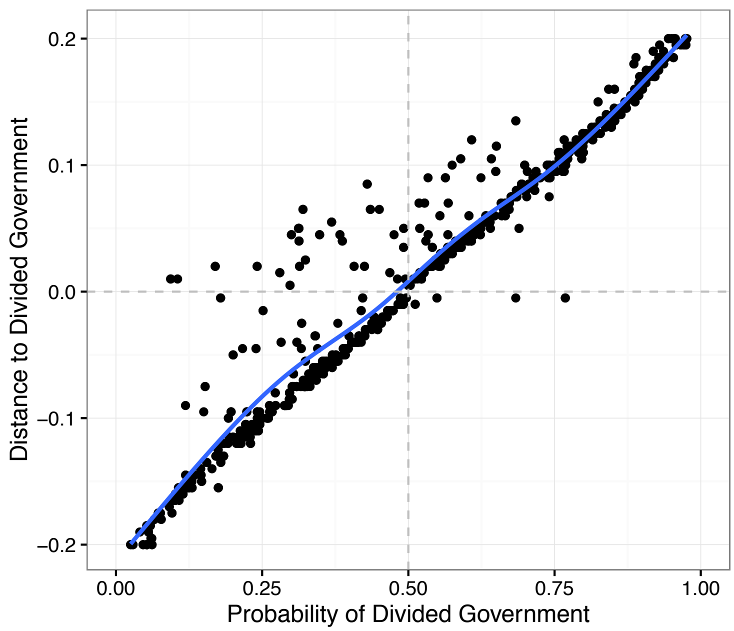
Although we use the distance to divided government as the forcing variable in our RDD analysis, we use the simulated probability of divided government to formulate the weights for our inverse probability weighting (IPW) strategy. We find that our simulated measure of the probability of divided government performs quite well, correctly predicting divided government in nearly 96% of cases. Figure A3 plots the percent of correctly predicted observations across the range of the

Table A1: The Effect of Divided Government on Covariates

Covariate	Optimal Bandwidth	Point Estimate	Standard Error	p-value
Government Shutdown	0.142	-0.048	0.074	0.513
Biennial Budgeting	0.221	0.046	0.058	0.429
Session Length	0.091	16.929	10.917	0.121
Supermajority Requirement	0.067	-0.063	0.053	0.236
Per-capita Income Change	0.106	1.065	0.698	0.127
Election Year	0.197	-0.008	0.063	0.903

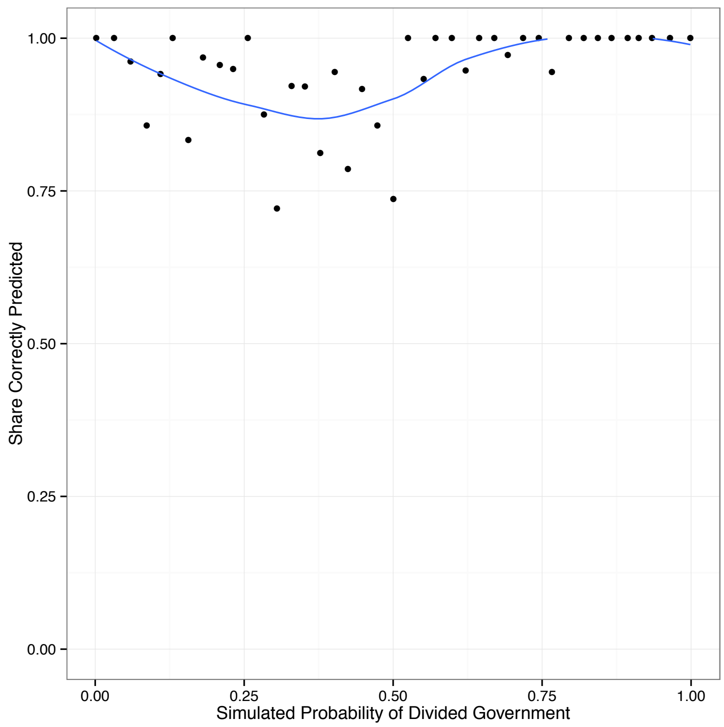
Note: Local linear regression models. Maximum of conventional and robust standard errors reported. Two-tailed test.

Figure A2: Probability of Divided Government & Distance to Divided Government



probability measure. In general, we do better at correctly categorizing states as our predicted probability of divided government moves away from 50%. Interestingly, when we err, we tend to do so by over-predicting unified government. We are unsure why this is the case. We do note, however, that incorrect predictions tend to occur close to the 50% threshold, i.e., where divided and unified government are about as likely to occur.

Figure A3: Simulated Probabilities & Divided Government



Note: This figure displays how accurately our simulated probabilities predict divided government across the values of the rating variable. The points plot the average of correctly predicted cases in bins of 2.5% of the rating variable.

C Alternative Distributions for Electoral Simulations

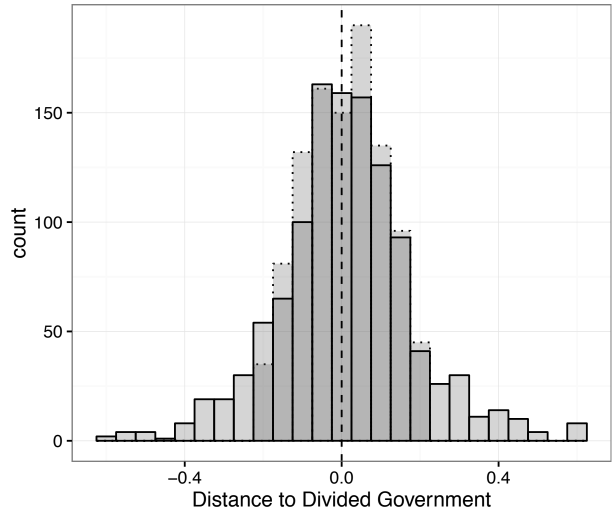
Throughout our simulation procedure, we rely on random draws from the uniform distribution to establish state- and district-level vote shocks. Recall that our goal is to find the smallest vote shock that likely would have generated a different configuration of party control of state government. Using the uniform distribution ensures that we have simulations that incorporate widely varying vote shocks across the range of possible shocks. To construct the forcing variable for our RDD, we choose the smallest vote shock that produces a different outcome in terms of divided government. As a result, we would expect use of a normal distribution with mean 0 would produce nearly identical values of the forcing variable.

Yet another possible approach would be to use actual historical vote shocks to generate the electoral shocks used in our simulation. For example, we could use the historical elections data to assemble a distribution of district-level shocks for each state and use random draws from each state's distribution to determine its state- and district-level vote shocks. Again, however, because our simulations are designed to find the smallest electoral shock that would have produced the opposite outcome in terms of divided government, we expect that this approach would also generate nearly equivalent values of the forcing variable. Indeed, when we replicate our simulation procedure (albeit for a smaller number of iterations) drawing electoral shocks from each state's distribution of historical electoral shocks, the resulting forcing variable is highly correlated with our main forcing variable ($\rho = 0.995$, $p < 0.001$). Figure A4 compares the distribution of our main forcing variable and the alternative measure determined using actual historic vote shocks.

D An Alternative to Simulations: A Centering Approach

As we discuss in Section 3.2, divided or unified government results from multiple elections, which creates complications in formulating a suitable assignment variable. In our main RDD analysis, our simulated measure of the distance to divided government serves as the forcing variable. However, we could take a different approach, using actual vote-share margins. Several earlier studies (e.g. Reardon and Robinson 2012; Wong, Steiner, and Cook 2013) have considered techniques

Figure A4: Comparison of Distributions of Forcing Variables



Note: The histogram displays the distribution of forcing variables using bins of 5%. Bins for the distribution of the forcing variable generated using historic shocks are outlined by a dashed line, while bins for the primary forcing variable are overlayed and outlined with a solid line. 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.

for RDDs in which treatment assignment is determined by multiple variables. Several applications investigate the effects of remedial education programs when students are assigned to treatment on the basis of multiple test scores, but we can also use this method to develop a forcing variable for divided government. Our approach is akin to what Wong et al. (2013, 109) refer to as "centering," a procedure in which multiple forcing variables are collapsed by choosing the value closest to the threshold that determines treatment assignment.[1]

To implement the centering or binding-score approach we need to determine the minimum vote-share margin that would have resulted in the opposite outcome in terms of divided government. That is, if a state experienced divided (unified) government in a given year, we need to collapse multiple assignment variables across legislative chambers and the executive branch to identify the smallest shift in vote-share that would have produced unified (divided) government. The assignment variable for the governor's party is simply the vote-share margin between the two or top two candidates. The analogous variables for the upper and lower house, however, are less

[1] Reardon and Robinson (2012) describe this approach as a "binding-score RD."

straightforward. Because seat share determines party control of each chamber, we need to incorporate the majority party's legislative seat margin to find the pivotal district election that would shift party control of a chamber. We focus on district-level elections that the minority party lost, sorting vote-share margins from smallest to largest (by chamber). For each chamber, we choose the margin that would change the outcome in the election for the pivotal seat.

For the sake of clarity we provide an example using Pennsylvania's 1982 state general election. The governorship, all 203 seats in the PA House of Representatives, and half of the 50 seats in the Senate were up for election. The state's 1982 elections jointly produced divided government, with Republican Governor Richard Thornburgh winning reelection to face a split legislature. Democrats won 103 seats to control the lower chamber, and Republicans held control of the upper chamber with 27 seats.

To determine our single forcing variable, we begin by examining the vote-share margins of party control for each institution. Table A2 shows the results of the gubernatorial election. Republican Richard Thornburgh won reelection with 2.72% more votes than his opponent (centered margin of 1.36%). Democrats had 2-seat margin in the lower chamber, so to find the vote share margin that would have given Republicans a majority, we need to find the second closest race lost by a Republican. Table A3 includes the two closest Democratic victories among the House elections. The Republican candidate in district 32 lost by a margin of 0.47%, so with a 0.235% (centered margin) shift in the two-party vote, Republicans would have won both of their narrowest losses to take control of the House. Analogously, if Democrats had won 3 additional Senate seats, they would have controlled the upper chamber. Table A4 lists the three closest Democratic Senate losses of the 1982 election. If 1.04% of the vote share shifted from Republicans to Democrats across all districts, Democrats would have won these three seats to take control of the Senate.

Table A2: 1982 PA Gubernatorial Election Results

Candidate		Vote Share	Margin	Centered Margin
Thornburgh	(R)	50.84%	+2.72%	+1.36%
Ertel	(D)	48.12%	−2.72%	−1.36%

Table A3: 1982 PA House Elections—closest Republican losses

	Margin Rank	District	Republican Margin	Republican Centered Margin
	1	89	−0.16%	−0.08%
•	2	32	−0.47%	−0.23%

Table A4: 1982 PA Senate Elections—closest Democratic losses

	Margin Rank	District	Democratic Margin	Democratic Centered Margin
	1	16	−0.28%	−0.14%
	2	30	−1.26%	−0.63%
•	3	44	−2.08%	−1.04%

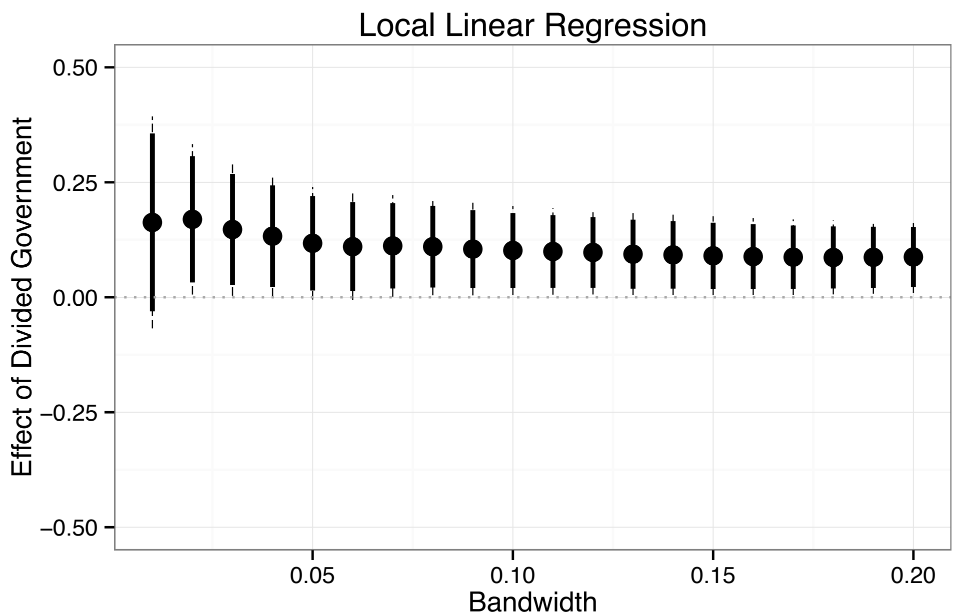
If Republicans had won control of the House or if Democrats had won both the gubernatorial election and a Senate majority, the result would have been unified government. Therefore, the collapsed forcing variable will be the smaller of either (1) the margin that would have switched party control of the House or (2) the larger of the governor's vote share margin and the margin that would have shifted party control of the Senate. The 1982 PA elections would have produced unified Republican party control if 0.235% of the vote had shifted to Republicans, and a 1.36% shift from Republicans to Democrats would have left Democrats in control of a unified state government. Therefore, the centered forcing variable for divided government is 0.235%.

Throughout the paper, our RDD specifications incorporate our simulated measure of the distance to divided government as the forcing variable. We take this approach for two main reasons. First, our simulation method also allows us to produce a measure of the probability of divided government which we use in our IPW analysis. Second, the centering option described above requires the assumption that shifts in legislative vote-shares would be uniform across districts while the simulations allow for variation across districts. Our approach is similar to methods developed by Folke (2014) and Fiva, Folke, and Sørensen (2013) to study the effects of party in proportional representation systems.[2] We did, however, replicate our main analysis with a vote-share forcing

[2] For a more thorough discussion of various legislative rating variables, including seat shares, vote shares and

variable constructed using the centering or binding-score method. The two assignment variables are highly correlated ($\rho = 0.956$), and we obtain similar results with either measure. Using local linear regression models that incorporate our non-simulated vote-share forcing variable, we estimate that divided government increases the probability of a late budget by 11.2% using the I & K optimal bandwidth (0.067). Figure A5 plots the point estimates and confidence intervals for the effect of divided government across multiple bandwidths, and we note that the pattern here is quite consistent with the results we present in Figure 5 of the paper.

Figure A5: The Effect of Divided Government—alternative vote-share forcing variable



The figure plots the effect of divided government across multiple bandwidths. The horizontal axis measures the bandwidth size, and the vertical axis measures the effect size. The 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 dashed lines indicate 95% confidence intervals.

E An Alternative Dependent Variable: Days Late

Here extend the primary analysis in our manuscript by considering the effect that divided government has on the length of delay. That is, conditional on a budget being late, is it later during periods of divided government than it is during periods of unified government. Perhaps

simulated distance measures, see Fiva, Folke, and Sørensen (2013) unsurprisingly, the answer is yes. Our models uncover a divided government effect of between 23 and 28 additional days. Stated somewhat differently, fiscal stalemate lasts, on average, three weeks longer during divided government.

Table A5: The Effect of Divided Government on Days Late

	Dependent variable:		
	Days Late		
	(1)	(2)	(3)
Divided government	28.844	33.703	23.398
	(20.681)	(16.737)	(11.962)
Distance to divided gov't	−308.557	−366.497	−131.588
	(667.787)	(365.875)	(130.239)
Distance to divided gov't * Divided government	450.954	290.532	144.963
	(849.947)	(441.917)	(169.457)
Constant	10.611	10.194	17.949
	(16.532)	(13.709)	(9.444)
Bandwidth	0.05	0.10	0.20
Observations within bandwidth	66	100	153
R^2	0.091	0.065	0.041
Adjusted R^2	0.047	0.036	0.022
Residual Std. Error	25.545 (df = 62)	29.041 (df = 96)	33.694 (df = 149)
F Statistic	2.074 (df = 3; 62)	2.219 (df = 3; 96)	2.133 (df = 3; 149)
Note:	Two-tailed test		

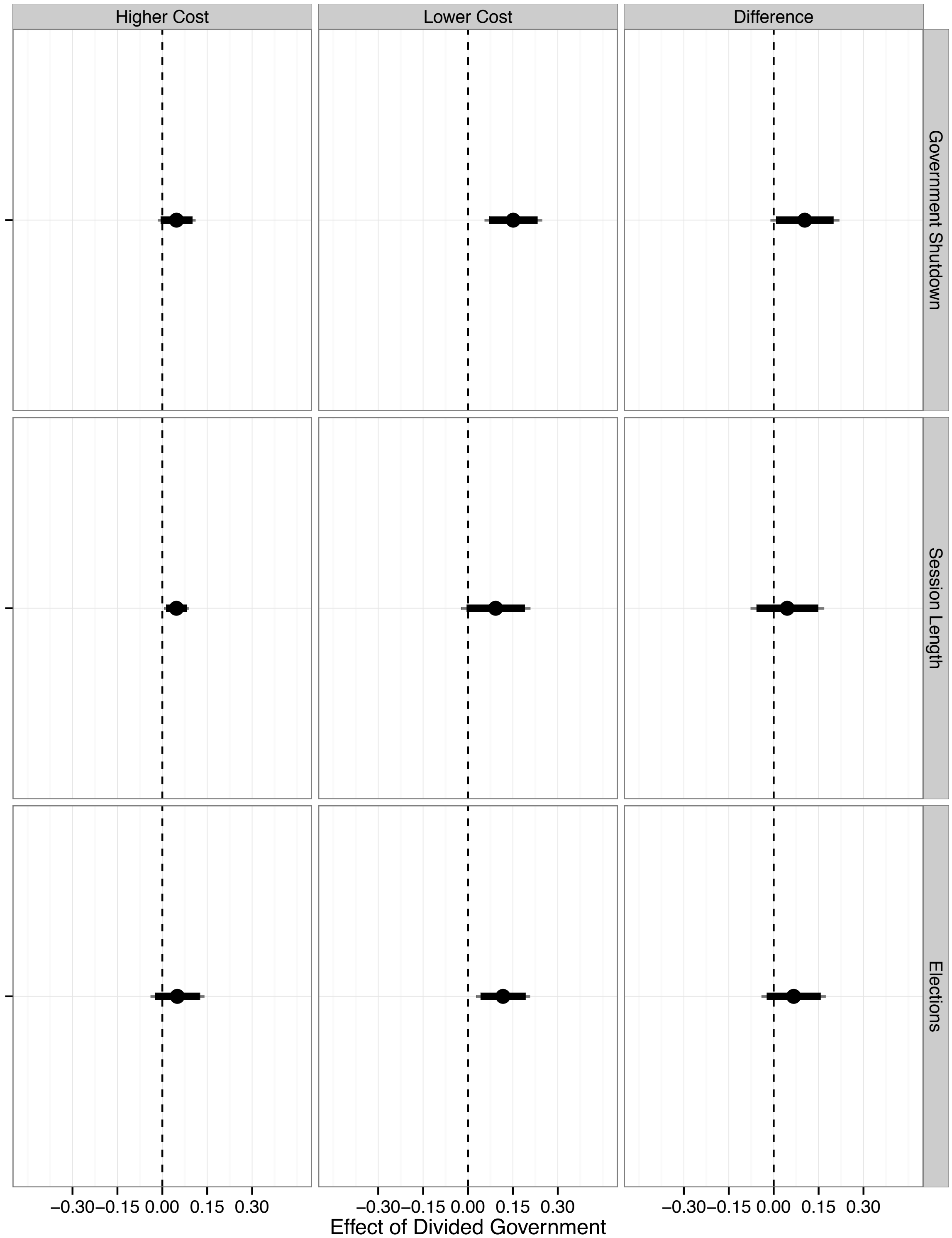
F Heterogeneous Effects: IPW & FE Analyses

In Section 3 of the paper, we use our RDD to investigate whether the effect of divided government on legislative delay varies with the political context. We find that divided government seems to have the greatest impact when the costs of impasse to politicians are low. Our analysis also yields some evidence that divided government may have a larger effect under high levels of polarization. Here, we replicate our analysis of heterogeneous treatment effects using both IPW and FE models.

In Figure A6, we present IPW estimates of the effect of divided government when the political and private costs of delay are high and low (these estimates are from an IPW model without state and year fixed effects, see Column 1 of Table 2). Unsurprisingly, these results are analogous to those presented above. When the costs of delay are high, the effect of divided government ranges from a relatively modest 4.5 to 5 percentage points. However, this effect is at least twice as large when costs are low. As was the case in our RDD analysis, these differences are not always statistically meaningful, though the difference for government shutdown is it significant at the 90% level. In states with a shutdown requirement, divided government results in a 4.7 percentage point increase in the probability of a late budget; in states without such a requirement, it results in 15.1 percentage point increase.

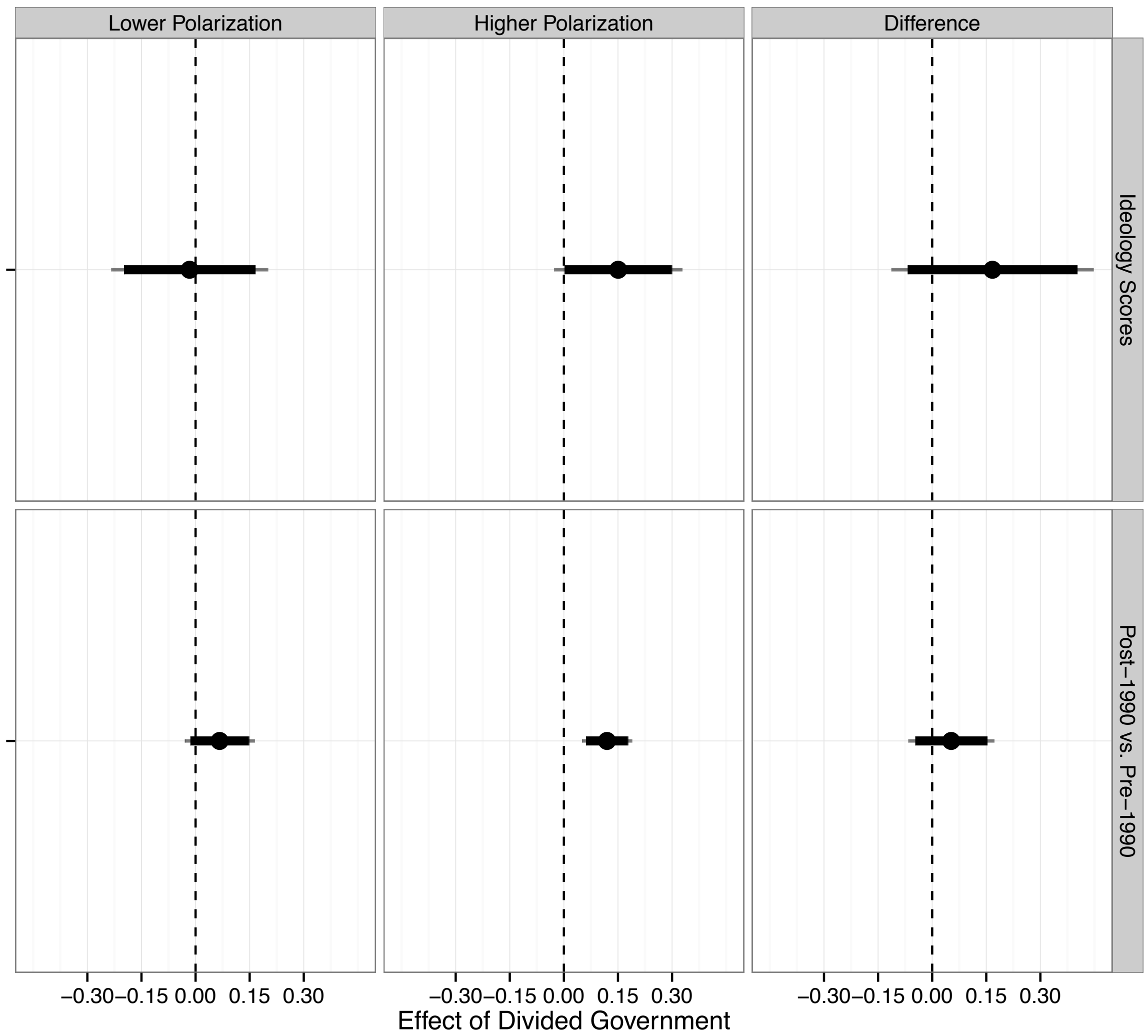
When we shift our focus to political polarization, we also find similar results using the IPW approach. In Figure A7, we again observe that effect of divided government appears to be greater when the level of polarization is high. When we use the Shor and McCarty ideology scores to measure polarization, our IPW model actually yields a negative, though small (less than 2 percentage points) and not statistically significant, effect of divided government under conditions of low polarization. This jumps to 15 percentage points during relatively high levels of polarization (though this difference is large, it fails to reach conventional levels of statistical significance). Likewise, during the first half of our time period (the lower polarization era), divided government increases the probability of fiscal delay by about 7 percentage points, jumping to 12 percentage points in the second half. Again, this difference is not statistically meaningful.

Figure A6: Heterogeneous Effects of Divided Government (IPW Results)



This figure summarizes the heterogeneous effects of divided government. The x-axis measures the effect size while the y-axis indicates the bandwidth. The black dots indicate point estimates from IPW models. Error bars illustrate 90% and 95% confidence intervals using 2-tailed tests.

Figure A7: Heterogeneous Effects of Divided Government (IPW Results)

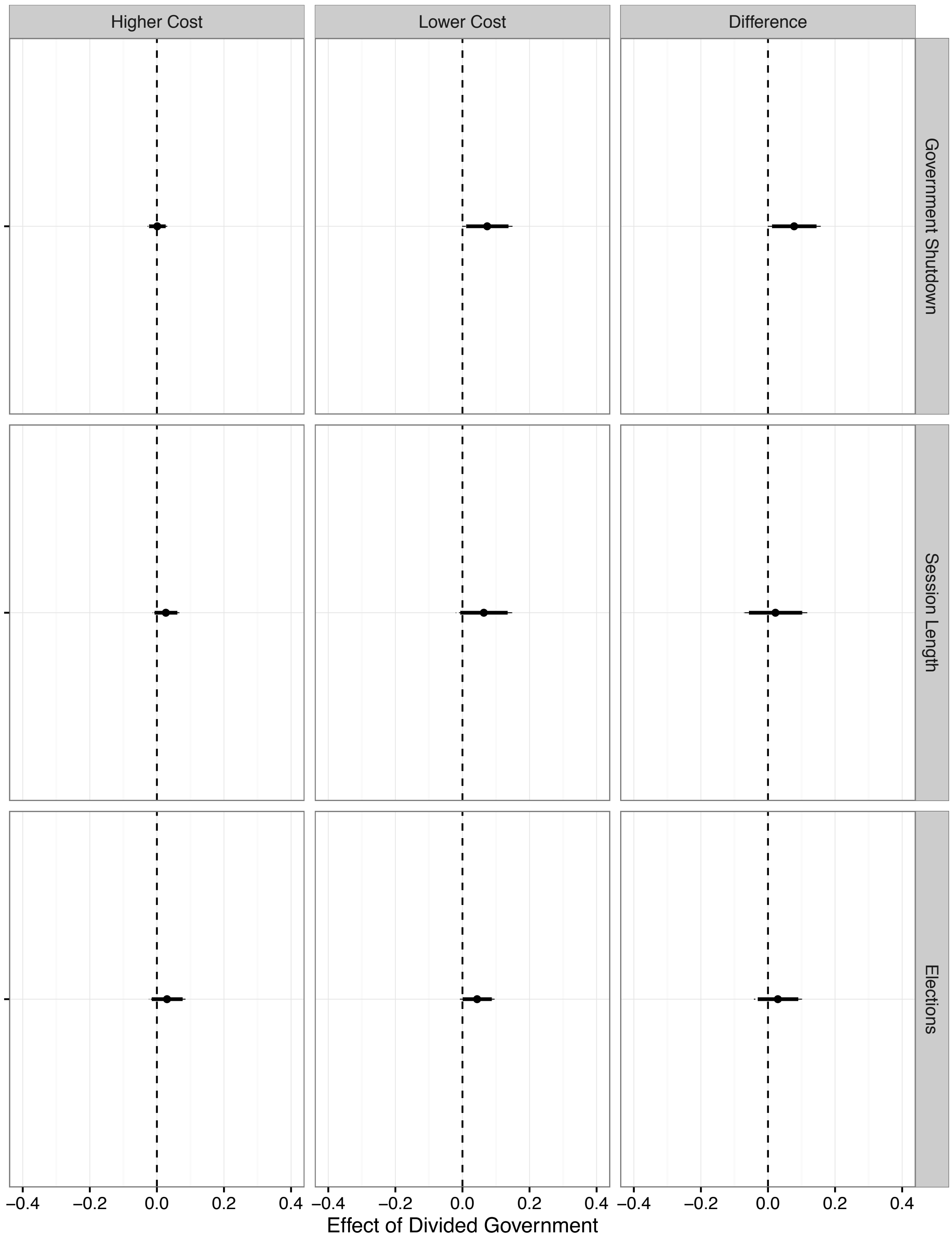


This figure summarizes the heterogeneous effects of divided government. The x-axis measures the effect size while the y-axis indicates the bandwidth. The black dots indicate point estimates from IPW models. Error bars illustrate 90% and 95% confidence intervals using 2-tailed tests.

Perhaps not surprisingly, the results of our FE models follow a familiar pattern. When we include state fixed effects, the estimates presented in A8 indicate that the effect of divided government on legislative delay tends to be greater when the cost of delay to lawmakers is low. As in our main FE analysis, we also note that the magnitude of the estimates is smaller across the board. Indeed, the effect of divided government on legislative delay ranges from close to 0 to about 3 percentage points when the costs of delay are high. When costs are low, however, divided government increases the likelihood of a late budget by about 4 to 8 percentage points. Though many of these differences are not statistically significant, the effect of divided government increases from nearly 0 in states with government shutdown rules to almost 8 percentage points in states without shutdown provisions, and this difference is statistically significant ($p = 0.051$).

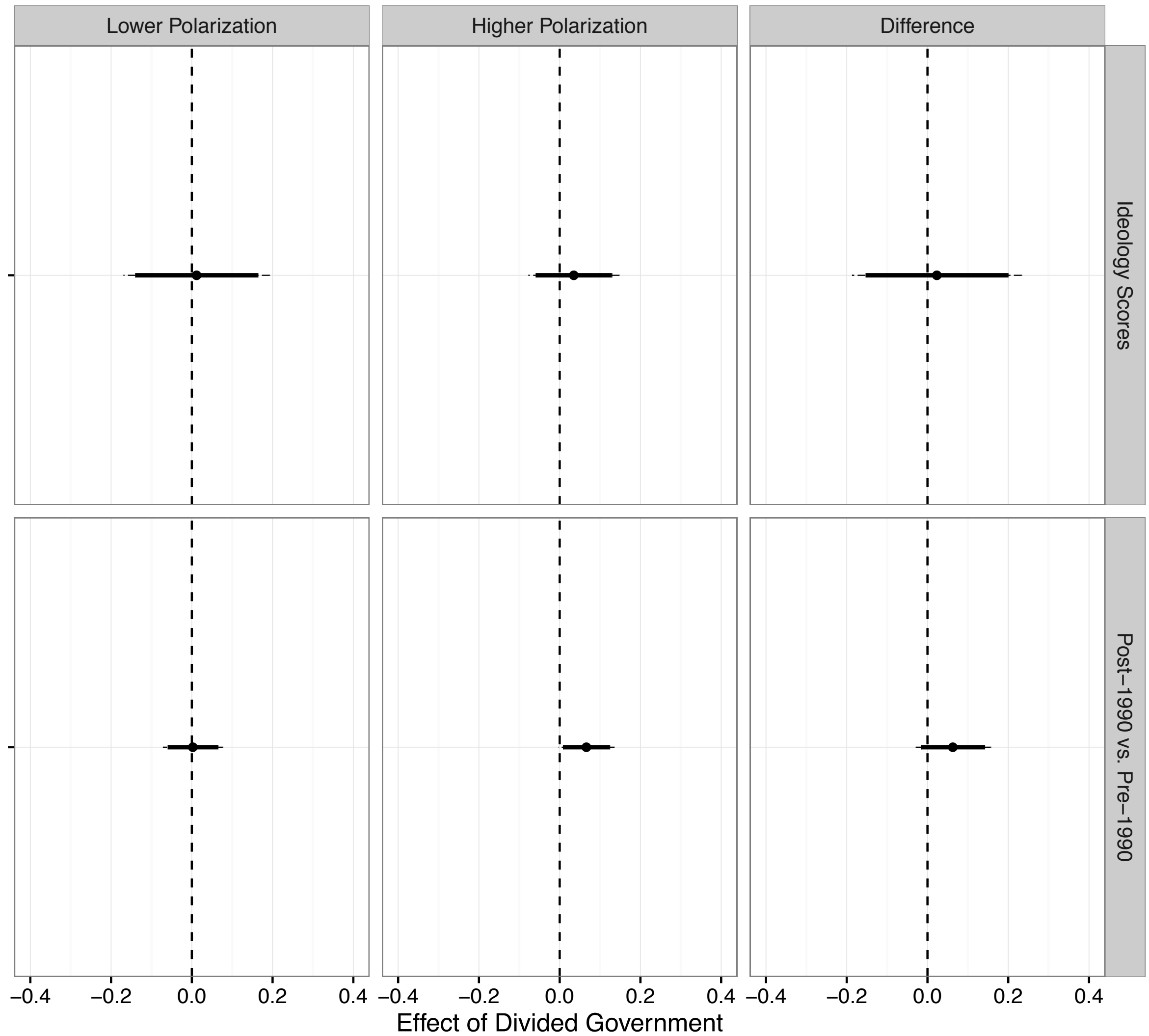
Moving on to consider polarization, the FE results displayed in Figure A9 are consistent with our the results of our RDD and IPW analyses—that is, the effect of divided government appears to be greater when polarization is relatively high. When we account for polarization using the 1990 cutoff, the effect of divided government is less than 1 percentage point during the earlier, lower-polarization time period compared to nearly 7 percentage points in the post-1990 era of higher polarization. If we employ ideology scores to gauge polarization, we also find a larger effect of divided government when polarization is higher. However, these differences are not statistically significant.

Figure A8: Heterogeneous Effects of Divided Government (FE Results)



This figure summarizes the heterogeneous effects of divided government. The x-axis measures the effect size while the y-axis indicates the bandwidth. The black dots indicate point estimates from models including state fixed effects. Error bars illustrate 90% and 95% confidence intervals using 2-tailed tests.

Figure A9: Heterogeneous Effects of Divided Government (FE Results)



This figure summarizes the heterogeneous effects of divided government. The x-axis measures the effect size while the y-axis indicates the bandwidth. The black dots indicate point estimates from models including state fixed effects. Error bars illustrate 90% and 95% confidence intervals using 2-tailed tests.

References

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

Fiva, Jon H, Olle Folke and Rune J. Sørensen. 2013. "The Power of Parties." CESifo Working Paper Series No. 4119. http://ssrn.com/abstract=2221882.

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

McCarty, Nolan, Keith T Poole and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. MIT Press.

Reardon, Sean F. and Joseph P. Robinson. 2012. "Regression Discontinuity Designs with Multiple Rating-score Variables." *Journal of Research on Educational Effectiveness* 5(1):83–104.

Shor, Boris and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(03):530–551.

Wong, Vivian C., Peter M. Steiner and Thomas D. Cook. 2013. "Analyzing Regression-discontinuity Designs with Multiple Assignment Variables a Comparative Study of Four Estimation Methods." *Journal of Educational and Behavioral Statistics* 38(2):107–141.