The impact of term-limit policies on the behavior of state lawmakers: new experimental findings from the Arkansas Senate*

Andrew Feher †
Department of Political Science
University of Michigan

Rocío Titunik‡
Department of Political Science
University of Michigan

April 30, 2015

Abstract

We analyze the impact of term-limit policies on the legislative output of state lawmakers. Several non-experimental studies have studied related questions, but inferences are complicated by the dynamic biases that may arise as a result of electoral defeat. We discuss and illustrate these methodological challenges with data from several states that adopted term limits, and show that paying close attention to research design clarifies the methodological obstacles. We then present an original experimental study that addresses many of these obstacles and is based on the random assignment of term length that occurs in the Arkansas Senate, which in turn induces the random assignment of term limits in 1997 and 2007. Across four measures of legislative output—bills introduced, bills passed, resolutions and abstention rates—we find no evidence of term limits effects. Since our experimental sample is small, we perform randomization-based inference, which is exact in finite samples. We also test the null hypothesis that term-limited lawmakers differ from their reelection-eligible counterparts, which allows us to assert that term limits decrease legislative output by at most a very small amount. Finally, we use bounds to address the fact that some senators in our original experimental sample are not observed.

*We thank Steve Cook, the Chief Council of the Arkansas State Senate, for providing details regarding the drawing of lots following censuses; Blake Potts at the Arkansas Legislative Digest for providing the roll-call data; the Roy Pierce Fellowship at the Center for Political Studies, University of Michigan, for financial support; and Jeremy Gelman, Marjorie Sarbaugh-Thompson, Chuck Shipan, and participants at the 2014 State Politics and Policy Conference for thoughtful comments and discussions.
†Ph.D. Candidate, Department of Political Science, University of Michigan, <afeher@umich.edu>.
‡Assistant Professor, Department of Political Science, University of Michigan, <titiunik@umich.edu>, http://www.umich.edu/~titiunik, Department of Political Science, University of Michigan, 5700 Haven Hall, 505 South State St., Ann Arbor, MI 48109.
1 Introduction

In 1990, voters in California, Colorado and Oklahoma approved initiatives limiting the number of terms that a lawmaker could serve in the state legislature, setting in motion what many scholars viewed as the most significant policy change in state government since the legislative modernization movement of the late 1960s (Mooney, 2009). Between 1992-1996, seventeen more states followed suit. As of April 2015, fifteen states jointly housing 37% of the total U.S. population still limit state legislators’ length of service.¹

Restricting the number of terms lawmakers can serve is an idea dating as far back as the 18th century. In the first Constitution of the United States, the Articles of Confederation, legislators were limited to no more than three out of every six years in office; and during the Continental Congress, Thomas Jefferson famously advocated for term limits to “prevent every danger which might arise to American freedom by continuing too long in office” (Jefferson, 1893, 61). But by the time the subject resurfaced in the 1990s, when nearly half of the states passed term limits initiatives, term limits policies were seen in a less positive light as the goal of increased turnover was weighed against its potential costs—in particular, against the possibility that removing reelection incentives would lead legislators to decrease their effort or adopt ‘out of step’ ideological positions.

In this article, we study the impact of legislative term limits on state lawmakers’ legislative output and participation. Establishing empirically whether these effects exist and measuring their magnitude is important for several reasons. First, we wish to understand whether rules that allow legislators to serve without the prospect of future electoral accountability result in systematic changes in legislative output. Since term limits are still being adopted and modified in many states, their effects on legislative behavior should be incorporated in the future design and evaluation of these policies. Second, the immediate effect of term limits

policies on the legislative output of individual lawmakers may have downstream consequences for the way in which policy is produced in the states. State governments have significant authority to formulate and implement policy in areas as varied as environmental protection, intrastate commerce and education (Gerber and Teske, 2000) and the state legislature, as one of the central state government actors, can have an impact on the nature of policy adoption and diffusion. For example, the institutional capacity and resources available to the legislature may affect whether states follow a pattern of bottom-up or top-down policy adoption (Shipan and Volden, 2006), and the reelection incentives facing state lawmakers may alter the temporal pattern of policy diffusion (Nicholson-Crotty, 2009).

Previous literature on the impact of term limits in state legislatures suggests that they have an impact on legislative behavior. For example, Carey et al. (2006) and Powell et al. (2007) find that term-limited state legislators devote less time to securing rewards for their district and helping constituents deal with government. In addition, Sarbaugh-Thompson et al. (2004) report that term-limited legislators turn their attention away from constituents and toward interest groups. The nonexperimental nature of these studies, however, complicates the causal interpretation of these findings.

We contribute in two ways to the study of the effect of term limit policies in state legislatures. First, we describe and document the methodological challenges that arise when this question is studied with non-experimental or quasi-experimental designs. In contrast to experimental designs, in natural experiments and non-experimental designs a valid comparison group may not readily identifiable (Sekhon and Titiunik, 2012). We show that this methodological challenge arises in non-experimental studies of term limits effects. Any study of the impact of term limits must deal with the challenge that politicians serving their last term in office are often systematically different from those whose electoral horizons are longer, which complicates the ability to attribute last-term behavior to the lack of electoral incentives. This phenomenon is most evident when the decision to retire is entirely under the control of the individual politicians—e.g., legislators may retire preemptively when they anticipate
that their poor past performance may result in a loss. But these inferential complications do not necessarily disappear when the occurrence of the last term is determined by an exogenous rule. As we discuss in detail below, while term limit laws prevent state legislators who wish to run indefinitely for reelection from doing, they fail to eliminate the dynamic biases that may arise as a result of electoral defeat.

These methodological obstacles are the motivation for our second contribution, which is the study of term limits effects using an experimental strategy in the Arkansas Senate. The Arkansas Constitution includes two features—the random assignment of senators’ term length in the first election after reapportionment and term limits—the combination of which results in the random assignment of state senators to lame-duck status. In particular, senators randomly assigned four-year terms in the first election after reapportionment see one less session in office than those randomly assigned two-year terms, which gives us the unique opportunity to examine two legislative sessions in 1997 and 2007 where the group of legislators assigned four-year lots is term-limited while the group assigned two-year lots is still eligible for reelection. Drawing on these experiments, we empirically test hypotheses regarding the effects of term limits on several measures of legislative participation and output.

Our empirical analysis addresses several statistical challenges. Since the Arkansas Senate has only thirty-five members, the overall sample size in our experiment, which pools two cohorts, is relatively small. To address this issue, we use randomization-based inference techniques that are exact in finite samples. As we explain below, in the randomization-inference framework, the distribution of the test-statistic under the null hypothesis of no effect is entirely determined by the treatment assignment, which allows us to test the hypothesis of no effect with an exact finite-sample p-value instead of relying on large-sample approximations that may be inadequate with our small sample size.

Since we are unable to reject the null hypothesis of no term limits effect and are interested in asserting that term limits do not alter legislative output, we also test the hypothesis that the average outcomes of term-limited lawmakers differ from the average output of reelection-
eligible lawmakers. In these tests, which are common in biomedical studies, the type I error rate is the probability of declaring that the two groups compared are equivalent when in fact they are not (Berger et al., 1996). Thus, in setting the type I error rate at a given level, say 5%, we control the probability of declaring that term limits have no effect when in fact they do. For every outcome, we calculate the minimum discrepancy between the term-limited and non-term-limited group that leads to a rejection of the null hypothesis that both groups are different. As we show, these tests of equivalence allow us to assert with 95% confidence that term limits do not have large impacts on legislative output and participation. In particular, at this level, we can rule out all negative term limit effects except for effects of very small size.

Finally, since we have some attrition in our experimental sample, we estimate bounds on the average term limits effects for every outcome under the assumption that those lawmakers whose outcomes we do not get to observe would have been systematically high or low, potentially affecting our conclusions (see, e.g., Manski, 2003). This analysis indicates that non-random attrition does not seem to be driving our conclusions.

After accounting for these challenges, we find no evidence that term-limited legislators reduce their effort by introducing or passing fewer bills, providing less constituency service as proxied by resolutions or abstaining at a higher rate on roll-call votes. Our findings are at odds with several observational analyses of state legislators’ behavior in their last terms in office, and have important policy implications for states with term limits and states considering their adoption.

The remainder of the paper is organized as follows. In the next section, we develop theoretical expectations about last-term effects on legislative behavior that take into account the specific institutional constraints of state legislatures. From there, we discuss the difficulties faced by observational studies of legislative last-term effects and propose an improved observational design. We then present the details of our experimental research design. Next, we present our results, followed by a section that presents robustness checks based on bounds.
to address the issue of attrition in the original experimental sample. We conclude in the last section. Additional results are presented in an online Supplemental Appendix.

2 Conceptual Framework: Legislative Behavior Absent Reelection Incentives

The most fundamental mechanism by which last-term effects are expected to arise is the removal of reelection incentives. Under an accountability model of representation, voters incorporate politicians’ past actions into their voting decisions and elections serve as an accountability mechanism that sanctions representatives’ behavior. In turn, the threat of punishment induces reelection-seeking politicians to behave in accordance with constituents’ preferences and expectations. Under this model, the logical consequence of adopting term limits is to induce undesirable legislative behavior or shirking, as the threat of punishment is removed and legislators have no incentive to please the electorate.\(^2\) Thus, if elections’ main role is to serve as an accountability mechanism, removing the possibility of running for reelection should result in systematic changes in legislative behavior (Fearon, 1999, p. 63).\(^3\)

Moreover, there may be mechanisms by which the removal of reelection incentives may result in lower legislative participation and output that are not directly related to the removal of electoral accountability. One such mechanism is the potential opportunity costs of seeking future employment. If legislators harbor some degree of progressive ambition and hope to secure their next occupation—higher office or otherwise—before their tenure comes to an end, lame-duck legislators face a trade-off: continue to actively participate in legislative activities or curb some of that legislative participation in order to invest attention in surveying their future employment options. Given the time commitments associated with casework and with

\(^2\)See Mansbridge (2009) for a discussion of the sanctioning model and Barro (1973) and Ferejohn (1986) for classical models on the control of politicians via reelection incentives.

\(^3\)We consider critiques to the accountability model (e.g., Fearon, 1999; Mansbridge, 2009) in the conclusion.
building the coalitions needed to successfully navigate bills through the legislature, we might expect those who are in their last term, and thus more likely to be in search of their next job, to reduce the effort they expend on constituency service and policymaking.

These two non-exclusive scenarios—shirking induced by the removal of reelection incentives and shirking induced by the opportunity costs of securing future employment—imply that term-limited legislators will have less time or incentives to meet with staff to help draft legislation, to learn about the kinds of bills their colleagues are sponsoring, to jockey for support in committee and on the floor to ensure passage of those bills that they do introduce, to allocate attention to casework and to attend roll-call votes. Accordingly, we can hypothesize that at the level of the individual legislator, term-limited members of the chamber will reduce their effort, (1) introducing fewer bills, (2) passing fewer bills, (3) performing less constituency service and (4) abstaining on a greater proportion of roll-call votes than non-term-limited members.

Legislative term limits vary enormously in their severity (see review by Sarbaugh-Thompson, 2010). Some states only impose bans on consecutive years of service, allowing legislators to cycle back and forth between both legislative chambers. This alters but does not completely remove reelection incentives, as legislators may plan to come back to their old district once they spend the mandatory number of years out of office. In contrast, some states impose lifetime bans on reelection, completely eliminating the possibility of ever running for the same seat and inducing a much more drastic reduction of electoral incentives.

Last-term effects will also depend on the degree of the legislature’s professionalism. Facing low pay, limited staff resources, poor advancement prospects and the absence of a reelection incentive, members serving in one of the twenty-four states with low-professionalization legislatures have few incentives to actively participate (Squire, 1988; Maestas, 2000). In these low-salary and limited-resource settings, state legislators must find time to negotiate the balance between outside careers, from which they derive their primary source of income, and their legislative obligations, leaving them with scant time to devote to legislative tasks. By
contrast, in professional legislatures, not only do staff subsidize the cost of policymaking, but these legislators earn salaries that permit them to devote all of their time to legislating. The incentives introduced by the removal of electoral accountability may therefore be amplified in less professional legislatures.

However, low professionalism could also attenuate last-term effects, in particular through mitigating the need to secure future employment. Less professionalized legislatures tend to conduct the preponderance of legislative business in regular sessions held in odd-numbered years. Thus, the trade-off between legislating and searching for future employment may never arise; instead, most legislators might be able to make the required three-month commitment every other year while maintaining an alternative source of employment on which they can continue to rely after term limits are imposed. Below, we use our experimental study to investigate these hypotheses empirically.

3 Research Designs to Study Last-Term Effects

A possible strategy to study term limits effects is to compare the outcomes of term-limited legislators to the outcomes of those legislators who can still run for reelection. This type of observational research design is a common choice (see, e.g., Carey et al., 1998, 2006; Powell et al., 2007; Sarbaugh-Thompson et al., 2004), and it is often the only one available. However, the two groups compared in this design may be systematically different and thus threaten the validity of the inferences. This may occur for at least two reasons.

First, legislators who are serving their last term because of term limit restrictions have by construction survived the highest possible number of elections that a legislator is allowed to contest before term limits come into effect. In contrast, legislators who are not yet term-limited are of two types: the “deportor” type, composed of legislators who will be defeated and will depart before term limits are binding, and the “survivor” type, composed of legislators who will win all elections and serve the highest possible number of terms under
the current term limit laws. The result is that, at any given point in time, the group of term-limited legislators is composed entirely of survivors, while the group of non-term-limited legislators is composed of both survivors and departors.

The differences between both groups will likely be systematic and related to the outcomes of interest. Since departors are candidates that will eventually be defeated, they are likely to be of lower average quality—i.e., less competent—than survivors, making the non-term-limited group of lower average quality than the term-limited group (since the latter is composed exclusively of survivors).\(^4\) Differential quality between both groups might lead to systematic differences in future performance for a number of reasons. For example, strong challengers may be deterred at higher rates in the term-limited group due to the higher quality of incumbents in this group. Naturally, it is not possible to distinguish both types in the group of legislators who are not yet term-limited, and thus these underlying differences cannot be “controlled for.”\(^5\)

Second, the group of term-limited legislators may differ systematically from the group of non-term-limited legislators due to the incumbency advantage and the phenomenon of strategic waiting by challengers. Term-limited legislators are, by definition, veteran incumbents. Thus, the incumbency status they enjoyed at the time of their last election could have translated into a less competitive election due to the advantages brought about by incumbency, including, possibly, a weaker challenger and increased name recognition. In contrast, depending on the specific term-limits restrictions, some or all legislators in the non-term-limited group will have been elected in open-seat races. These races will tend to be more competitive than incumbent races, which may translate into higher pressure to display good

---

\(^4\)The result that more competent legislators are more likely to survive reelection is standard in agency models of electoral selection (see, for example, Alt et al., 2011, and references therein).

\(^5\)This difficulty applies not only to studies of term limits, but also to other types of studies such as those that compare retiring and non-retiring Congress members in lame-duck sessions (see, e.g, Carey, 1998). Note, however, that the situation might be inverted in lame-duck congressional sessions: since there are no term limits in the U.S. Congress, those retiring might have served less terms on average than those who are returning to the chamber, which might result in returning members being of higher average quality than departing members. This is more likely to occur, for example, when retirements are due to anticipated bad performance than when they occur because the member has reached retirement age.
performance. Moreover, differences in the degree of competitiveness between term-limited and non-term-limited legislators could arise because challengers might prefer to wait until the seat becomes open instead of challenging an incumbent in his last eligible election. The result would be higher rates of uncontested races and as a result less competition among term-limited legislators.

In sum, either because of intrinsic differences in candidate quality due to the different number of terms survived, differences in incumbency status, or strategic waiting by challengers, and likely due to the combination of all of these factors, non-experimental comparisons of term-limited versus non-term-limited legislators may not lead to valid estimates of term-limits effects. To provide evidence of these inferential problems, we compare the electoral fortunes of term-limited and non-term-limited legislators in their most recent election in eight state legislatures—four with consecutive service bans and four with lifetime service bans. We pool observations from several election cycles, comparing cross-sectionally the mean vote shares and the proportion of uncontested races between term-limited and non-term-limited legislators in their most recent election.

The first four rows in Table 1 report results from the lower and upper chambers in the Colorado, Ohio, Arizona and South Dakota legislatures, all of which have consecutive service bans. We see that there is a substantively large and statistically significant difference in vote share in all four states: term-limited (TL) legislators in these states enjoy vote shares between 4-5 percentage points higher than non-term-limited (NTL) legislators. In two states, term-limited legislators are also significantly less likely to face a challenger. The remaining four rows in Table 1 report results from the lower chamber in the Arkansas legislature and the lower and upper chambers in the California, Michigan and Oregon legislatures, all of which have lifetime service bans. Again, term-limited representatives have significantly higher vote shares than non-term-limited representatives, though the size of the differences is more variable, ranging from roughly 2 to 10 percentage points depending on the state. Moreover, with the exception of Arkansas’ lower chamber, term-limited legislators are not significantly
Table 1: Uncontested Rates and Difference-in-means for Vote Share Between Term-Limited and Non-Term-Limited State Legislators

<table>
<thead>
<tr>
<th>States with Consecutive Bans</th>
<th>Vote Share</th>
<th>Uncontested Rates</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean NTL</td>
<td>Mean TL</td>
<td>p-val</td>
</tr>
<tr>
<td>Arizona</td>
<td>49.45</td>
<td>53.74</td>
<td>.09</td>
</tr>
<tr>
<td>Colorado</td>
<td>67.41</td>
<td>71.51</td>
<td>.01</td>
</tr>
<tr>
<td>Ohio</td>
<td>68.19</td>
<td>72.60</td>
<td>.00</td>
</tr>
<tr>
<td>S. Dakota</td>
<td>47.20</td>
<td>52.02</td>
<td>.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>States with Lifetime Bans</th>
<th>Vote Share</th>
<th>Uncontested Rates</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean NTL</td>
<td>Mean TL</td>
<td>p-val</td>
</tr>
<tr>
<td>Arkansas (House)</td>
<td>83.45</td>
<td>93.92</td>
<td>.00</td>
</tr>
<tr>
<td>California</td>
<td>65.73</td>
<td>67.92</td>
<td>.01</td>
</tr>
<tr>
<td>Michigan</td>
<td>66.71</td>
<td>68.31</td>
<td>.06</td>
</tr>
<tr>
<td>Oregon</td>
<td>66.54</td>
<td>71.85</td>
<td>.02</td>
</tr>
</tbody>
</table>

Note: The election data come from the State Legislative Election Returns (1967-2010) ICPSR #34397 dataset. Column labeled ‘Mean TL’ reports the mean outcome for term-limited senators, and column labeled ‘Mean NTL’ reports the mean outcome for non-term-limited senators. The columns labeled ‘p-val’ report p-values from two-tailed t-tests of the null hypothesis that means are equal. Column labeled N reports the number of observations used in each row.

Overall, the results show that, on average, term-limited legislators are more electorally successful than reelection-eligible legislators, a result consistent with the aforementioned possible confounders. Moreover, these vote share differences are not solely explained by the higher rate of uncontested races among term-limited legislators. As we show in Section A2 of the Supplemental Appendix, the differences in vote shares persist and become even more pronounced when uncontested races are excluded from the analysis.

Below we present an experimental design that addresses most of these challenges, but first we consider a non-experimental research design that may mitigate these inferential problems in cases where an experimental design is not available. We propose a research strategy that restricts the comparison group to only those non-term-limited legislators who have successfully won at least one reelection bid. There are several reasons why eliminating freshman legislators from the comparison group might alleviate some of the biases. First, if
most low-quality incumbents are defeated in their first reelection, restricting the comparison
group to non-freshman incumbents may eliminate weaker incumbents and thus increase the
average candidate closer to the average quality among survivors in the term-limited group.
The systematic differences between the groups could also be alleviated by the fact that
most candidates eliminated from the non-term-limited group will be incumbents elected in
(relatively) competitive open seats. The veteran incumbents who stay in the non-term-
limited group after freshmen are eliminated are thus likely to have been elected, on average,
in an environment more similar to the environment faced by non-term-limited incumbents.

Table 2: Uncontested Rates and Difference-in-means for Vote Share Between Term-Limited
and Non-Term-Limited State Legislators, Excluding Freshman Legislators

<table>
<thead>
<tr>
<th>States with Consecutive Bans</th>
<th>Vote Share</th>
<th></th>
<th>Uncontested Race</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean NTL</td>
<td>Mean TL</td>
<td>p-val</td>
<td>Mean NTL</td>
<td>Mean TL</td>
</tr>
<tr>
<td>Arizona</td>
<td>54.32</td>
<td>53.74</td>
<td>.83</td>
<td>14.65</td>
<td>16.30</td>
</tr>
<tr>
<td>Colorado (House)</td>
<td>71.29</td>
<td>74.01</td>
<td>.21</td>
<td>19.53</td>
<td>25.00</td>
</tr>
<tr>
<td>Ohio (House)</td>
<td>71.66</td>
<td>73.14</td>
<td>.32</td>
<td>15.42</td>
<td>19.08</td>
</tr>
<tr>
<td>S. Dakota</td>
<td>52.15</td>
<td>52.02</td>
<td>.96</td>
<td>12.56</td>
<td>14.02</td>
</tr>
<tr>
<td>States with Lifetime Bans</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arkansas (House)</td>
<td>91.19</td>
<td>93.92</td>
<td>.00</td>
<td>75.29</td>
<td>82.40</td>
</tr>
<tr>
<td>California (House)</td>
<td>68.04</td>
<td>68.15</td>
<td>.93</td>
<td>5.79</td>
<td>3.66</td>
</tr>
<tr>
<td>Michigan (House)</td>
<td>69.63</td>
<td>68.98</td>
<td>.54</td>
<td>2.08</td>
<td>1.59</td>
</tr>
<tr>
<td>Oregon (House)</td>
<td>72.08</td>
<td>70.35</td>
<td>.59</td>
<td>20.63</td>
<td>16.67</td>
</tr>
</tbody>
</table>

Note: The election data come from the State Legislative Election Returns (1967-2010) ICPSR #34397
dataset. Column labeled ‘Mean TL’ reports the mean outcome for legislators who are term-limited and
column labeled ‘Mean Non-TL’ reports the mean outcome for legislators who are not term-limited. The
columns labeled ‘p-val’ report p-values from two-tailed t-tests of the null hypothesis that means are equal.
Column labeled N reports the number of observations used in each row.

Consistent with our expectations, once we exclude freshmen from the group of non-
term-limited legislators, uncontested rates and vote shares between the two groups become
considerably more similar. Table 2 presents uncontested rates and vote shares from the
same eight states in Table 1 after removing freshman legislators. Because the state senates
in Colorado, Ohio, California, Michigan and Oregon only permit two four-year terms, all
non-term-limited senators are freshman, so we cannot apply this research design for these chambers. For this reason, in these states we only examine lower chamber election results. By contrast, in Arizona and South Dakota legislators in both chambers can serve up to four two-year terms, so we continue to examine both the lower and upper chambers for these two state legislatures. With the exception of Arkansas’ lower chamber, there are no distinguishable differences in uncontested rates and vote shares between term-limited and reelection-eligible legislators. This suggests that observational studies of term-limits effects might be improved by excluding freshman legislators from the comparison group.

4 A Research Design Based on Random Assignment

Motivated by the methodological challenges in observational studies of term limits just discussed, we use an experimental design that relies on the random assignment of term length, which in turn induces the random assignment of term limits later in the decade. This design avoids many of the inferential problems mentioned above, as the group of term-limited and non-term-limited legislators are on average identical at baseline due to the initial random assignment.

Our research design is based on the random assignment of term length in the Arkansas Senate. Arkansas senators normally serve a term of four years and their terms are staggered, with (roughly) half of the 35 senate seats up for election every two years. However, Article 8, Section 6 of the state’s constitution mandates that, in the first election following a decennial census and the corresponding redrawing of district boundaries, all 35 seats must be up for election. Since the simultaneous election of all 35 seats breaks the staggering of terms, term lengths are randomly assigned to return the chamber to staggered terms.

Specifically, Section 6, Amendment 23, of the Arkansas Constitution instructs senate seats to be randomly divided into two classes of size 17 and 18 after each reapportionment. The

---

6The motivation for this provision is to ensure that all sitting senators have been elected by their new constituencies.
pattern of term length differs by class: senators elected to a seat in the class of size 18 serve a
two-year term immediately following reapportionment and a four-year term thereafter, while
senators elected to a seat in the other class serve two successive four-year terms immediately
following redistricting and a two-year term at the end of the decade. Senators draw lots at
the beginning of the first legislative session immediately after redistricting to determine the
composition of each class of seats. This design, and similar designs in Illinois and Texas, was
used by Titiunik (2015) to study the effects of term length on legislative behavior.

In November 1992, 60 percent of Arkansas voters supported Amendment 73, a term limits
initiative that was among the most stringent in the country. This amendment limited state
representatives’ service to a lifetime maximum of three two-year terms and state senators’
service to a lifetime maximum of two four-year terms. A very important element of our
research design is that two-year terms do not count against the two-term limit—only four-
year terms do.

Our goal is to measure outcomes for term-limited and non-term-limited senators during
the same legislative session, to avoid conflating time differences and genuine last-term effects.
For this reason, we study two cohorts of senators for whom term limits become effective
during the same legislative session: those first elected or reelected in 1992, and those first
elected in 2000 or 2002. All Arkansas senators elected in November 1992, whether elected
for the first time or reelected, served their last period either in 1996-2000 or 1998-2002 (if
they did not retire or lose sooner). Senators elected in 1992 for a two-year term could run for
reelection in 1994 for a four-year term (1994-1998), and again in 1998 for another four-year
term (1998-2002), because the first two-year term did not count towards term limits. In
contrast, those elected in 1992 for a four-year term could run for reelection only once in 1996
to serve a second four-year term between 1996 and 2000. In other words, since 1992 is the
“baseline” year when term limits are adopted, regardless of how many times senators in this

7Throughout, we refer to the length of terms by an interval from an even year to another even year, such
as 1998-2002. The first year in the interval indicates the year when the election took place, and the last year
in the interval the last year of the term. For example, 1998-2002 refers to the term, served between January
1999 and December 2002, for which a senator was elected in November 1998.
cohort had been reelected prior to 1992, they would all serve their last allowed term at the same time, except for the 2-year discrepancy induced by the staggering.

The situation for later cohorts is different, because as some senators lose or retire before the maximum allowed number of terms, the newly elected senators’ last allowed terms occur at different points in time. If a few new senators were entering every year, it would be hard to study an additional cohort, as everyone would be term-limited at different times, invalidating our design. Luckily, there are only six senators who are elected for the first time between 1994 and 1998, with the remaining 29 first elected in either 2000 or 2002. These 29 senators constitute the second cohort in our analysis—since two-year terms do not count toward term limits, a first election in 2000 or 2002 leads to a final term during either 2006-2010 (if a four-year term is drawn in 2002) or 2008-2012 (if a two-year term is drawn in 2002).

Figures 1(a) and 1(b) illustrate the sequence senators experience based on whether they draw a two-year or a four-year term in the 1990s and 2000s, respectively. As the figures show, there are two legislative sessions in the Arkansas General Assembly, the 81st in 1997 and the 86th in 2007, during which senators randomly assigned four-year terms following reapportionment are ineligible to run for reelection or *lame ducks*, while those assigned two-year terms are eligible for their last reelection in the following election.
Consider two senators who entered the chamber in either 2000 or 2002 and won their races in 2002, one assigned a two-year term and one assigned a four-year term. Because a two-year term does not count toward the two-term lifetime limit, the senator assigned to serve a two-year term will stand for reelection in November 2004 and again in November 2008. By contrast, a state senator assigned a four-year lot in 2002 is already on the term-limit clock and will only stand for reelection one more time in November 2006. This in turn
makes for a legislative session in 2007 (the 86th) where senators assigned four-year lots in
2002 are lame-ducks while senators assigned two-year lots in 2002 still face an election in
2008. The sequence is analogous in the 1990s. For analysis, we pool both cohorts, totaling
64 senators (35 from 1992 cohort, 29 from 2000/2002 cohort), and study outcomes of interest
during the 81st and 86th regular sessions.

Similarly to observational studies, our design does require the Stable Unit Treatment
Value Assumption (SUTVA) in order to yield effects that can be interpreted as the effects of
term limits. When SUTVA holds, the outcome of every experimental unit is solely affected
by the treatment received by that unit, regardless of the treatment status assigned to the
rest of the units participating in the experiment (see, e.g., Rubin, 1990; Bowers et al., 2013).
In our research design, SUTVA requires that a legislator who is term-limited behave in the
same way regardless of how many other legislators in the chamber are term-limited. This
would restrict scenarios where, for example, non-term-limited legislators let term-limited
legislators have a larger share of those resources that have a fixed budget (e.g. floor time) to
help them take actions that will position them favorably in their quest for higher political
office. However, given that term-limited legislators are not returning to the chamber, these
agreements might be difficult to sustain in equilibrium (see Muthoo and Shepsle, 2010).
Moreover, this kind of strategic coordination may be less likely to occur for outcomes that
are not directly constrained by the actions of others (e.g. bill introductions). The fact that
our results are consistent across abstention rates and bill introductions thus alleviates our
concerns about possible SUTVA violations.

Validity of Experimental Research Design

We now provide evidence regarding the validity of the experimental research design just
described. Under random assignment of term length, all predetermined characteristics at

8Due to the aforementioned confusion surrounding passage of Amendment 73, lots were drawn after the
79th session in October 1993 instead of at the beginning of the session. In the post-2000 reapportionment, by
contrast, lots were drawn in December 2002, after the election but before the start of the legislative session.
the senator level are identical in expectation between senators assigned a two-year term (henceforth ‘2-year senators’) and senators assigned a four-year term (henceforth ‘4-year senators’) following reapportionment. Thus, if we observed significant dissimilarities between our two samples, the validity of the randomization might be called into question.

We present the results from balance tests from our full sample of senators described above who drew lots in October 1993 and December 2002. Since the total sample size is 64, we tested the null hypothesis that there is no treatment effect on predetermined covariates using randomization inference instead of parametric tests—see Rosenbaum (2010), §2, for an introduction to randomization inference methods. We chose the difference-in-means between the term-limited and non-term-limited groups as a test-statistic, and used the random assignment of state senators to either group to calculate the exact randomization distribution of this test-statistic under the null hypothesis of no treatment effect for any senator. Since this test is based on the known randomization of the treatment, it has the advantage that its distribution is exact even in small samples. In contrast, a t-test would require the validity of large-sample approximations, which may not be appropriate given our low sample size.⁹

As shown in the first four columns in Panel A of Table 3, both groups have similar means across seven pretreatment covariates, including vote share obtained in the previous election, party, and race. And using our randomization-based tests, we fail to reject the null hypothesis of no treatment effects for these covariates, as is expected in a successfully implemented experiment (minimum p-value across all covariates is 0.18). Given our relatively low sample size, however, our failure to reject the null hypothesis could be driven by a lack of power. For this reason, we also test the hypothesis that the term-limited and non-term-

---

⁹Since the enumeration of all possible realizations of the treatment assignment is not feasible, we based our tests on 10,000 simulations—where each simulation takes one treatment assignment at random from all possible treatment assignments. Since the random assignment of terms was done separately for each cohort, in our simulations we separately drew the treatment assignment of each cohort, and then pooled both cohorts to compute the difference in means between our pooled treatment and control groups. This implementation respects the way in which the randomization was actually performed. However, analyzing all observations jointly ignoring that the two cohorts were separately assigned produces qualitatively similar results.
limited groups are different.

Letting $\mu_{TL}$ be the mean in the term-limited group and $\mu_{NTL}$ the mean in non-term-limited group, we test the null hypothesis that the discrepancy or dissimilarity between both means is larger than a positive number $\delta$, that is, we test $H_0^\delta: |\mu_{TL} - \mu_{NTL}| > \delta$. These tests, sometimes referred to as equivalence tests, are commonly used in medical studies to establish bioequivalence between generic and brand-name drugs (Berger et al., 1996). Whereas in the common balance tests reported above the null hypothesis is that there is no difference in the legislative output between term-limited and non-term-limited legislators, with equivalence tests we make the null hypothesis that the groups are different and only reject it when there is sufficient evidence that the two groups are similar. Whether the $H_0^\delta$ is rejected naturally depends on the value of $\delta$. We report the minimum value of $\delta$ for which $H_0^\delta$ is rejected at 5% level. To reject $H_0^\delta$ for a given $\delta_*$ means that we can assert with 95% confident that legislative output in the term-limited group differs from the legislative output in the reelection-eligible group by at most $\delta_*$. Thus, if $\delta_*$ is small, we can assert that the groups are similar.

We report the results from equivalence tests in the last two columns of Panel A in Table 3. We report the minimum value of $\delta$ for which $H_0^\delta$ is rejected at 5% in two ways: in raw units and in standard deviation units—i.e., the quantity $\delta_*$ in raw units divided by each variable’s standard deviation in the pooled sample. For example, the first row shows that we can be 95% confident that the difference between the vote share obtained by term-limited senators and the vote share obtained by reelection-eligible senators is no greater than 11.03 percentage points in either direction, an absolute difference that represents 0.8 pooled standard deviations. Given that the average vote share is so high in both groups—between 85% and 90%, a difference of at most 11 percentage points in either direction would still imply that senators in both groups were elected by a very large margin. We reach similar conclusions for the other covariates, although the minimum $\delta$ corresponding to the Democrat and Black variable, 0.32 and 0.21, respectively, are fairly high, a result we attribute to the low sample sizes. Taking the results from both the tests of no effect and the equivalence

<table>
<thead>
<tr>
<th></th>
<th>Means TL</th>
<th>Means NTL</th>
<th>Difference</th>
<th>p-value</th>
<th>Min δ for which $H_0^δ$ is rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Raw units</td>
</tr>
<tr>
<td>Vote Share</td>
<td>89.27</td>
<td>86.8</td>
<td>2.46</td>
<td>0.6</td>
<td>11.03</td>
</tr>
<tr>
<td>Married</td>
<td>0.88</td>
<td>0.88</td>
<td>0</td>
<td>1</td>
<td>0.17</td>
</tr>
<tr>
<td>Male</td>
<td>0.94</td>
<td>0.81</td>
<td>0.12</td>
<td>0.23</td>
<td>0.28</td>
</tr>
<tr>
<td>Democrat</td>
<td>0.91</td>
<td>0.75</td>
<td>0.16</td>
<td>0.18</td>
<td>0.32</td>
</tr>
<tr>
<td>Black</td>
<td>0.12</td>
<td>0.06</td>
<td>0.06</td>
<td>0.67</td>
<td>0.21</td>
</tr>
<tr>
<td>Attorney</td>
<td>0.25</td>
<td>0.34</td>
<td>-0.09</td>
<td>0.55</td>
<td>0.26</td>
</tr>
<tr>
<td>Age</td>
<td>50.66</td>
<td>52.78</td>
<td>-2.12</td>
<td>0.45</td>
<td>6.98</td>
</tr>
</tbody>
</table>

Sample size: 32 (TL group) and 32 (NTL group)

Note: ‘TL’ refers to term-limited senators (assigned 4-year lot in 1992 or 2002), and ‘NTL’ refers to non-term-limited senators (assigned 2-year lot in 1992 or 2002). The test of no effect reports randomization-based p-values corresponding to the sharp null hypothesis that the treatment has no effect for any unit using the difference-in-means as test-statistic. In tests of the hypothesis $H_0^δ$ reported in the last two columns, $µ_TL$ refers to the mean outcome among term-limited senators and $µ_{NTL}$ refers to the mean outcome among non-term-limited senators; these tests are also randomization-based, assuming a constant treatment effect model. SD units are expressed in terms of the pooled standard deviation across all senators.

A final issue that we address is sample attrition. By 1997 and 2007, the years when the 81st and 86st Legislative Sessions begin, 15 senators in our sample of 64 had left the chamber: 12 senators left between 1993 and 1997 and 3 senators left between 2002 and 2007. This leaves us with a remaining sample of 49 senators, whom we call “compliers”, and brings some complications to the design. Any time attrition occurs in an experimental setting it raises concerns that the remaining subjects no longer represent a random sample of the original experimental sample, which in turn would lead to invalid inferences (see, e.g., Gerber and Green, 2012). This would occur in our case if a senator’s defeat or retirement before term limits become binding is affected by the term length assigned after reapportionment—i.e., by our treatment of interest.
In this and further sections, we present arguments and evidence that suggest we can make meaningful inferences despite this methodological challenge. First, we note that attrition levels are comparable across treatment and control groups. As shown in the last row of Table 3, our initial sample size is 32 senators in each group, and after attrition there are 26 senators in the term-limited group and 23 senators in the non-term-limited group, samples sizes that are entirely consistent with a 1/2 probability of assignment to each group.\textsuperscript{10} Moreover, this balance in attrition levels is also seen when we consider each legislative session individually.\textsuperscript{11}

Note also that, in addition to the initial randomization—which allows us to ensure comparability at baseline—a crucial aspect of our design is that both groups of senators have survived the same number of elections (one) when the outcomes are observed. As a result, the attrition that results from electoral defeat in the first reelection is likely to affect both groups equally, and the composition of both groups in terms of departors and survivors is thus likely to be similar at the moment when outcomes are measured. Moreover, this composition is equal (on average) at baseline due to the initial randomization.

Nonetheless, and despite losing roughly the same number of senators in each group, there could still be differences in the type of senators who drop out. For example, if the senators who drop out in the term-limited group are more productive on average than the senators who drop out in the control group, we might observe that the term-limited group has lower legislative output than the non-term-limited group, but we would be mistaken to attribute this difference to last-term effects. The assumption that there is no endogenous attrition is inherently unobservable, so we cannot directly provide evidence that it does not happen in our case. For this reason, in the results section we use bounds to explore whether our conclusions are sensitive to the possibility of endogenous attrition. In general, we find that

\textsuperscript{10}The null hypothesis that the true probability of success is equal to 0.5 in 49 trials of a Bernoulli experiment cannot be rejected with 23 successes (p-value 0.7754). And the difference in compliance rates by group (23/32 and 26/32) are statistically indistinguishable (t-test p-value is 0.3840).

\textsuperscript{11}In the 1990s cohort, 8 senators assigned 2-year terms and 4 senators assigned 4-year terms drop out of the sample before 1997 (null hypothesis that true probability of success is 0.5 in 12 Bernoulli trials is not rejected, p-value 0.3877), and in the 2000s cohort 1 senator assigned a 2-year term and 2 senators assigned 4-year terms drop out of the sample before 2007.
even allowing for endogenous attrition term limits do not seem to lead to lower productivity.

5 Last Term Effects in the Arkansas Senate

We now study our main question of interest, whether term-limited state senators engage in less legislative activities than their reelection-eligible counterparts. To do so, we examine five dependent variables at the individual level: the number of bills introduced, the number of bills passed, the abstention rate on roll-call votes and the number of resolutions, which we use as a proxy for constituency service. While we would prefer a more conventional measure of constituency service (e.g. number of district staff, trips back to the district), they are not readily available. Instead, we use data on the resolutions that state senators file during the legislative session. As is typically the case with constituency service, these resolutions are devoid of ideological content; examples include recognizing the achievements of a citizen within their district or congratulating a local high school for its athletic accomplishments. We thus use the number of resolutions as an imperfect proxy for constituency service.

Theoretical Expectations

The framework developed above allows us to state specific expectations for the effects of term limits. In Squire’s (2007) index of legislative professionalism, Arkansas ranked 39th in 1996 and 41st in 2003, and in the National Conference of State Legislature’s (NCSL) Red-White-Blue trifurcation, Arkansas is considered a “White,” or hybrid, legislature based on its intermediate-sized staff and salary, as legislators do not earn enough to make a living without having other sources of income (National Conference of State Legislatures, 2009). The General Assembly holds its regular session in odd-numbered years, meeting for approximately sixty days, and holds what are variously known as fiscal sessions or extraordinary sessions in even-numbered years.

During the 1980s, before term limits were adopted, legislative turnover in Arkansas was
low: approximately half of the house seats and two-thirds of the senate seats were occupied by veteran legislators during this decade (Sarbaugh-Thompson, 2010, Table 1). Given this low turnover before term limits and stringent limits on length of service that followed, the effects of term limits on the Arkansas Senate should be higher than in most other states, where the institutional change induced by more lenient term limit policies did not represent such a drastic change (Sarbaugh-Thompson, 2010, p. 202). In addition, applying the framework discussed above, the relatively low professionalization of the Arkansas State Legislature is likely to exacerbate these last-term effects even more.

On the other hand, the need to secure future employment is likely not a major factor in this setting. In Arkansas, English and Weberg (2007, 148) note, “the part-time nature of the Assembly provide[s] members with ample opportunity to earn a living in their primary vocation while also serving their constituencies as lawmakers.” Occasionally, these legislatures convene for special shorter sessions in even-numbered election years, which Arkansas staff describe as “uneventful” and “pro forma” affairs that have historically involved simply rubber-stamping budgets. This means that the opportunity cost of seeking future employment is not likely to induce additional changes in legislative behavior.

In sum, although the need for future employment is not likely to be a major factor in inducing last-term effects, the initial low levels of legislative turnover, the stringent term limit restrictions adopted in the early 1990s, and its low level of professionalization, make the Arkansas State Legislature an environment where the effects of removing reelection incentives, if real, should be detected.

**Results**

Our analysis of the outcome variables mirrors the analysis of covariates reported above. We start by testing the null hypothesis that the effect of term limits is zero for every senator with randomization inference, using the difference-in-means between term-limited senators

---

12 Personal communication with Arkansas senate staff.
(those randomly assigned four-year lots) and non-term-limited senators (those randomly assigned two-year lots) as the test-statistic. We also provide randomization-based confidence intervals for the term limits effect, assuming that this effect is constant.\textsuperscript{13} As before, we pool observations across the two cohorts to maximize the number of observations. We report separate analyses for the 1997 (81st) and 2007 (86th) sessions in Section A4 of the Supplemental Appendix, which lead to comparable results.

Table 4 provides these inference results for our four dependent variables, and Figures 2 and 3 display the entire distributions using box-plots. We start by describing bill outcomes. The average number of bills introduced is remarkably similar between both groups, with term-limited senators introducing about 1.5 more bills than the non-term limited senators. Using this test-statistic, we are far from rejecting the sharp null hypothesis of no effect (p-value 0.65), while the 95% confidence interval ranges from -4.77 to 7.39 bills. A similar result is observed regarding the number of bills passed, with term-limited senators again marginally outperforming their non-term-limited colleagues, passing approximately 1 more bill during the session. Again, though, we are unable to reject that term limits have no effect, and the confidence interval is shorter than in the case of bill introductions. Turning to our measure of constituency service, the third row of Table 4 shows that term-limited senators average a little over two resolutions while non-term-limited senators file just over 1.5 resolutions during the legislative session. This difference is small and, again, we fail to reject the null hypothesis of no effect. Moreover, the 95% confidence interval covers mostly positive effects. Figures 2(a), 2(b) and 2(c) show that, looking at the entire distributions, there is no evidence that the term-limited produces less legislative output according to these bill measures.

An additional measure of participatory shirking involves abstention rates on roll-call

\textsuperscript{13}We calculate this confidence interval by inverting a hypothesis test in a constant treatment effect model. That is, letting $Y_{i,TL}$ denote the potential outcome of senator $i$ under term limits and $Y_{i,NTL}$ denote the potential outcome of senator $i$ in the absence of term limits, we constructed a 95% confidence interval by adopting the model that $Y_{i,TL} = Y_{i,NTL} + \tau$, testing the null hypothesis that $\tau = \tau_0$ for all possible values of $\tau_0$, and keeping the hypotheses that we failed to reject at 5% level.
Table 4: Effects of Term Limits on Legislative Behavior in Arkansas Senate, pooling 81st (1997-1998) and 86th (2007-2008) Legislative Sessions

<table>
<thead>
<tr>
<th>Means</th>
<th>Test of no effect</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>TL</td>
<td>NTL</td>
<td>Difference</td>
</tr>
<tr>
<td>Bills introduced</td>
<td>22.88</td>
<td>21.48</td>
</tr>
<tr>
<td>Bills passed</td>
<td>14.58</td>
<td>13.35</td>
</tr>
<tr>
<td>Resolutions</td>
<td>2.23</td>
<td>1.57</td>
</tr>
<tr>
<td>Bills Cosponsored</td>
<td>36.69</td>
<td>40.85</td>
</tr>
<tr>
<td>Abstention rate</td>
<td>2.36</td>
<td>1.14</td>
</tr>
<tr>
<td>Abstention rate w/o Faris</td>
<td>1.21</td>
<td>1.14</td>
</tr>
</tbody>
</table>

Sample size: 23 (TL group) and 26 (NTL group)

Note: ‘TL’ refers to term-limited senators (assigned 4-year lot in 1992 or 2002), and ‘NTL’ refers to non-term-limited senators (assigned 2-year lot in 1992 or 2002). The test of no effect reports randomization-based p-values corresponding to the sharp null hypothesis that the treatment has no effect for any unit using the difference-in-means as test-statistic. Confidence interval calculated by inverting randomization-based hypothesis tests in a constant treatment effect model. Calculations for abstention rates excluding Steve Faris (in the TL group) use a total sample size of 48.

votes. The term-limited senators abstain on approximately one percent more roll-calls than non-term-limited senators, leading to yet another failure to reject the sharp null hypothesis. The confidence interval is not symmetric around zero, ranging from a small negative effect of -0.65 to a larger positive effect of 3.35. This disparity in the mean abstention rates between groups which translates into a confidence interval that is shifted to the right of zero is driven entirely by one senator in the treatment group, Steve Faris, who missed approximately 31 percent of roll-call votes during the 2007 session. As the fourth and fifth rows of Table 4 and Figures 3(a) and 3(b) indicate, when we drop this senator from the analysis, the mean differences in abstentions between both groups vanish almost entirely and the confidence interval becomes much shorter and approximately symmetric about zero.
Figure 2: Term limit effects bill introduction, passage, and symbolic bills in Arkansas Senate—81st (1997-1998) and 86th (2007-2008) Legislative Sessions

(a) Bills Introduced

(b) Bills Passed

(c) Resolutions Filed
Figure 3: Term limit effects on abstention rates in Arkansas Senate—81st (1997-1998) and 86th (2007-2008) Legislative Sessions

In sum, our randomization-based inferences suggest that there is no evidence of term limits effects on four measures of legislative output and participation. For three of our outcomes, the mean point estimates indicate that term-limited senators are more active, not less, than their non-term-limited counterparts, and the box-plots of the entire distributions equally show that there is no evidence of shirking. However, failing to reject the null hypothesis of no effect does not necessarily mean that we can be confident in asserting that the outcomes in the two groups are equivalent. This is true in every application, and it is a more pressing concern in our case due to the low sample size, which affects our ability to detect true differences.

We therefore examine the impact of term limits using tests of equivalence, as we did for the covariates. As explained above, in these tests, our null hypothesis is that the legislative
output in the term-limited group is sufficiently different from the legislative output in the reelection-eligible group. The first two columns of Table 5 report how large of a disparity in legislative output, in either a positive or negative direction, would lead to a rejection of the null hypothesis of nonequivalence. Based on the absolute values in the first column, we can say with 95% confidence that the effect of term limits is at most about 7 bill introductions, 5 bills passed, a 3 percent abstention rate and 1.5 resolutions. In other words, these are the smallest differences between the groups that we can reject with a 5% level test.

Since many previous studies have found that term limits lead to lower effort or output—sometimes referred to as legislative shirking—we are particularly interested in whether we can rule out large negative term limits effects. To explore this issue, we test the hypothesis the term-limited and non-term-limited groups differ by a negative amount, that is, we test the null hypothesis of shirking \( H_0^{S} : \mu_{TL} - \mu_{NTL} < -\delta \), for a positive \( \delta \). Note that \( \mu_{TL} - \mu_{NTL} < 0 \) implies shirking only for the bill and resolution outcomes, but not for our abstention rates, since a positive effect of term limits on abstentions is a shirking effect. For this reason, when we report this one-sided test for abstention outcomes, we transform the outcome to be one minus the abstention rate, so that \( \mu_{TL} - \mu_{NTL} < 0 \) means that term-limited senators are voting less than their non-term-limited counterparts.

The results, which we report in columns 3 and 4 of Table 5, suggest that we can rule out large negative effects of term limits on legislative output and participation, as we can assert with 95% confidence that non-term-limited senators, at most, introduce three more bills, see two more bills passed, and file one-quarter more resolutions than their term-limited counterparts, effects that range from 0.2 to 0.4 standard deviations. In terms of abstentions, when Steve Faris is excluded, we can assert that the average rate of non-absent votes among non-term-limited senators is at most 0.74 percentage points higher than among term-limited senators, an effect of 0.8 standard deviations but small in absolute terms.

In sum, our experimental results from Arkansas fail to reject the null hypothesis that term limits have no effect on the measures of legislative output and participation we report,
and allow us to rule out with 95% confidence that term limits have large or even moderate negative effects.

Table 5: Tests of equivalence and negative effects (shirking), pooling 81st (1997-1998) and 86th (2007-2008) Legislative Sessions

<table>
<thead>
<tr>
<th>Test of equivalence at 95%</th>
<th>Test of shirking at 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0^k:</td>
<td>\mu_{TL} - \mu_{NTL}</td>
</tr>
<tr>
<td><strong>Min $\delta$ for which $H_0^k$ is rejected</strong></td>
<td><strong>Min $\delta$ for which $H_0^{k,S}$ is rejected</strong></td>
</tr>
<tr>
<td>RAW UNITS</td>
<td>SD UNITS</td>
</tr>
<tr>
<td>Bills introduced</td>
<td>7.4</td>
</tr>
<tr>
<td>Bills passed</td>
<td>5.52</td>
</tr>
<tr>
<td>Abstention rate$^a$</td>
<td>3.36</td>
</tr>
<tr>
<td>Abstention rate$^a$ w/o Faris</td>
<td>0.83</td>
</tr>
<tr>
<td>Resolutions</td>
<td>1.66</td>
</tr>
</tbody>
</table>

Sample size: 23 (TL group) and 26 (NTL group)

Notes: ‘TL’ refers to term-limited senators (assigned 4-year lot in 1992 or 2002), and ‘NTL’ refers to non-term-limited senators (assigned 2-year lot in 1992 or 2002); $\mu_{TL}$ refers to the mean outcome among TL senators and $\mu_{NTL}$ refers to the mean outcome among NTL senators. Tests of the hypotheses $H_0^k$ and $H_0^{k,S}$ are performed using randomization inference, assuming a constant treatment effect model; SD units are expressed in terms of the pooled standard deviation across all senators. Calculations for abstention rates excluding Steve Faris (in the TL group) use a total sample size of 48. $^a$In the last two columns where the test of $H_0^{k,S}: \mu_{TL} - \mu_{NTL} < -\delta$ is reported, the outcome used is one minus the abstention rate.

Robustness Check: Bounds

The previous section treated attrition as random. In this section, we explore whether our results survive patterns of retirement or defeat that are correlated with the initial assignment of term length. To address attrition in our experimental samples, we estimate upper and lower bounds on the average treatment effect following Manski (2003). In calculating the upper bound on the average treatment effect, we set the outcome values of those senators initially assigned to the treatment group (i.e., 4-year term length group) who were not present in the legislature during the 1997 and 2007 sessions equal to the 75th percentile of each of our four measures of legislative behavior in our pooled complier sample, and the missing outcome values of those senators initially assigned to the control group (i.e., 2-year
term length group) equal to the 25th percentile value. To calculate the lower bound on
the average treatment effect, we do the opposite, setting missing outcomes in the treatment
group at the 25th percentile value of the observed outcomes, and missing outcomes in the
control group at the 75th percentile value.

We believe this to be a plausible scenario to test the robustness of the results presented
above. The bounds calculated under this scenario essentially recompute the average treat-
ment effect assuming that the pattern of attrition is severely correlated with the initial term
length assignment. Our lower bound assumes that all missing outcomes in the treatment
group, if observed, would have been low (i.e., equal to the 25th percentile of the observed
outcomes) while all the missing outcomes in the control group would have been high (i.e.,
equal to the 75th percentile of the observed outcomes). Analogously, our upper bound as-
sumes that all missing outcomes in the treatment group, if observed, would have been high
(i.e., equal to the 75th percentile of the observed sample) while all the missing outcomes
in the control group would have been low (i.e., equal to the 25th percentile of the observed
sample).

Table 6 reports the bounds on the average treatment effect of term limits on legislative
behavior across the four outcomes. The columns ‘ATE Lower Bound’ and ‘ATE Upper
Bound’ report, respectively, the estimated lower and upper bounds of the difference-in-means
between the term-limited and non-term-limited groups. Since our intention is to establish
if the no-shirking results reported in the previous section are robust, we focus on the lower
bound for bill and resolution outcomes and on the upper bound for abstention rates. The
results in Table 6 show that for bill introductions, bill passage, resolutions and abstention
rates, even a severely endogenous pattern of attrition would result in small shirking or
negative term-limits effects. For example, for the number of bills introduced, assuming that
all missing term-limited senators would have introduced just 12 bills (25th percentile) while
all missing non-term-limited senators would have introduced 29 bills (75th percentile) would
result in a lower bound for the difference-in-means of just -2.75 bills, showing that even under

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>ATE Lower Bound</th>
<th>ATE Upper Bound</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bills introduced</td>
<td>-2.75</td>
<td>5.22</td>
<td>[-7.14, 9.49]</td>
</tr>
<tr>
<td>Bills passed</td>
<td>-1.78</td>
<td>3.84</td>
<td>[-4.82, 6.75]</td>
</tr>
<tr>
<td>Resolutions</td>
<td>0.03</td>
<td>0.97</td>
<td>[-0.64, 1.65]</td>
</tr>
<tr>
<td>Abstention rate</td>
<td>0.65</td>
<td>1.37</td>
<td>[-0.61, 3.18]</td>
</tr>
<tr>
<td>Abstention rate w/o Steve Faris</td>
<td>-0.27</td>
<td>0.43</td>
<td>[-0.76, 0.92]</td>
</tr>
</tbody>
</table>

Sample size: 32 (TL group) and 32 (NTL group)

Note: Columns labeled ‘ATE Lower Bound’ and ‘ATE Upper Bound’ report, respectively, the estimated lower and upper bounds of the difference-in-means between the term-limited (TL) and non-term-limited senator (NTL) groups. Upper bounds sets missing treated outcomes to 75th percentile of observed treated outcome and missing control outcomes to 25th percentile of observed control outcomes. Lower bound is analogous, using 25th percentile for missing treated and 75th percentile for missing control. Confidence intervals calculated with bootstrapping. Calculations for abstention rates excluding Steve Faris (in the TL group) use a total sample size of 63.

severely endogenous attrition we could rule out large last-term effects on this outcome.

A similar pattern is observed for bills passed, resolutions and abstention rates. The lower bound on the last-term effects on bill passage is just -1.78 bills, and this is assuming that missing term-limited senators would have passed just 7 bills while missing non-term-limited senators would have passed 19. The lower bound on resolutions is positive at 0.03, ruling out shirking. And the upper bound on last-term effects for abstention rates excluding Steve Faris is 0.43%, less than half of a percentage point, assuming that the missing abstention rates among term-limited senators would have been 1.68% while the abstention rate among missing non-term-limited senators would have been 0.16%, about ten times smaller—as shown, including senator Faris increases this bound to 1.37. The 95% confidence intervals on the estimated bounds interval are naturally consistent with larger negative effects, which is expected given the variability that stems from our low sample size. In sum, these results show that when it comes to abstention rates, bill introductions, bill passage and resolutions, our finding that there are no last-term effects on participatory shirking seems robust to endogenous attrition.
6 Conclusion

We have examined how the adoption of term limits affects legislative behavior, specifically the extent to which term-limited legislators produce less legislation and abstain at higher rates than their reelection-eligible counterparts. While other scholars have studied the effects of term limits in the U.S. states, ours is the first to experimentally examine the question. Our experimental design overcomes several of the static and dynamic methodological challenges that the nonrandom assignment of legislators to term-limited status has presented previous scholars. Leveraging two natural experiments in the Arkansas Senate that lead to the random assignment of term-limited status, we find no evidence that legislators slack off when the electoral connection is severed, placing our findings in contrast to previous studies of term limits in the American states.

Although our results are necessarily limited in scope because they cover only one state, the fact that we fail to see even small differences in the participatory measures we examine might cast some doubt on whether, in the United States, the essential role of elections is to sanction representatives and facilitate accountability. Several authors, (e.g., Fearon, 1999; Mansbridge, 2009), have advocated for a selection model of representation, where elections are seen primarily as mechanisms to select representatives that are self-motivated to act in the best interest of voters even in the absence of monitoring and sanctioning. Our results are consistent with this view, and suggest that an often overlooked cost of adopting term limits is to reduce the pool of politicians that can act as representatives.

Our results also have implications for public policy. The decision of citizens across nearly half of the U.S. states to limit the number of terms their representatives could spend in office promised to deliver sweeping changes in both the operation of state legislatures and in the behavior of state legislators. Critics of the initiative process, and of term limits in particular, argued that legislators would have carte blanche to act irresponsibly. Those fears of voters magnifying agency problems have not come to fruition, at least not in the Arkansas Senate.

Much remains to be learned about the dynamics of term limits in particular and the dy-
namics of last-term effects in general, and additional empirical work is needed to ensure that our conclusions hold for non-participatory outcomes and are generalizable beyond Arkansas. However, we believe that our findings may carry some implications for other states. As mentioned above, Arkansas’ lifetime term limits are among the most stringent in the country, suggesting that the change in the incentives facing legislators as they begin to serve their last term is likely larger in Arkansas than in most other states. If even in Arkansas term limits show no effect on legislative participatory outcomes, it is reasonable to expect that similar null effects would be found in other states where the removal of electoral accountability is neither so complete nor severe.

On the other hand, the scarce legislative resources associated with the low level of professionalization of the Arkansas Legislature might partly mediate the observed effects, in which case our results would not be immediately applicable to states with highly professional legislatures, where resources are abundant. But even in this case, our findings might carry implications for the six other low-professionalization or “dead-end” legislatures that currently have term limits, which comprise almost half of the total number of states where term limits are currently in effect.¹⁴

¹⁴In addition to Arkansas, the other dead-end legislatures include Arizona, Louisiana, Maine, Montana, Nebraska and South Dakota (Maestas, 2000).
References


