Supplementary Materials for "The Constraining Power of the Purse: Executive Discretion and Legislative Appropriations"

June 7, 2018

A Appropriations Committee Report Examples

The reports by appropriations committees that accompany bills to the House and Senate floors are important documents, treated by agencies as near-binding law. They instruct agencies on how the committee expects funds to be expended, place restrictions on their use for certain reasons, outline requirements for agencies in the coming fiscal year, and provide justifications for particular funding decisions (Fenno 1966; Kiewiet and McCubbins 1991; Schick 2008).

Common types of constraints included in committee reports (based on our reading) include:

- Stress the importance of particular programmatic goals relative to others in order to organize the agenda for an agency
- Reporting requirements: specify contents of reports on status of programs and spending
- Notification of spending
- Deadlines
- Directions on how to spend money
- Require justification from the agency for allocating funds for a particular purpose
- Gives instructions as to the particular focus a program should have when obligating funds
- Limitations on expenditures
- Requirements to make certain policy documents available to the public
- Requests for guidelines and certain procedures for implementation and enforcement
- Require the continuation of certain programs
- Requirements for the specification of duties for certain offices
- Requirement to submit a spending plan before obligation of in the next year's budget justification

- Provides a mission or objective for a particular program or office as the committee understands it
- Requirements to conduct studies analyzing program and spending effects.
- Require certain staffing levels through a full-time equivalent floor, ceiling, or target
- Instructions for the inclusion of certain information in next year's budget request
- Goals the agency should pursue for a particular program
- Require the agencies to develop and report plans to reduce costs in given program areas
- Specifies purpose of particular projects the committee wants to fund (most common in Energy and Water bills)

A.1 Example 1– Environmental Protection Agency

Below, we reproduce an except from House Report 112-589 describing appropriations for the Environmental Protection Agency (EPA). The report contains detailed information about how the appropriated funds are to be spent and specific information that the subcommittee requires to oversee the agency's activities. We argue that this type of language, which is typical of appropriations committee reports, can plausibly be considered constraints on agency language and that the length of these reports is positively correlated to the amount of constraint the committee intends to place on the agency.

Hydraulic Fracturing- In 2010, the Committee urged EPA to research whether there is a relationship between hydraulic fracturing and drinking water. The Committee understands EPA has incorporated a review of environmental justice impacts into this study, which the Committee finds to be outside the scope of the 2010 language and an inappropriate use of funds. No funds have been provided in the bill to research environmental justice impacts related to hydraulic fracturing, and EPA shall discontinue the use of any resources that may have been diverted to this subactivity. The Committee directs the Agency to release the study's findings with respect to whether there is a relationship between hydraulic fracturing and drinking water following appropriate public comment as directed in House Report 112-151 and peer review.

Integrated Risk Information System (IRIS).–The Committee strongly supports the goals of EPA's IRIS Program and believes a transparent, robust, and reproducible approach for synthesizing scientific information is an important element of influential Federal scientific assessment programs. However, it has become increasingly clear that fundamental improvements in the policies and practices of the IRIS program are necessary to ensure that the assessments developed are firmly based on up-to-date scientific knowledge, meet the highest of standards of scientific inquiry, and are evaluated in accordance with acceptable scientific approaches. Therefore, building from the directives in the fiscal year 2012 Interior, Environment, and Related Agencies conference report, the Committee directs the Agency to take the following actions:

(a) For draft and final IRIS assessments released in fiscal year 2013, the Agency shall include documentation describing how the Chapter 7 recommendations of the National Academy of Sciences (NAS) have been implemented or addressed, including an explanation for why certain recommendations were not incorporated.

(b) The Agency shall issue a progress report to the House and Senate Committees on Appropriations and relevant Congressional authorizing committees no later than March 1, 2013, describing the IRIS Program's implementation of the National Research Council's Chapter 7 recommendations.

(c) Accordingly, the Committee directs EPA to re-evaluate, using acrylonitrile and other relevant assessments as case studies, the methods previously used to evaluate and interpret the body of available scientific data, including the weight-of-evidence approach, and include in the report called for in section (b) a chapter on whether there are scientifically more appropriate methods to assess, synthesize and draw conclusions regarding likely human health effects associated with likely exposures to the substances.

A.2 Example 2 – The Marine Mammal Commission

Here, we present a pair of examples from reports pertaining to the Marine Mammal Commission, specifically in the House reports for the Departments of Commerce, Justice, State, the Judiciary and related agencies. The first report is House Report 110-919 for FY 2009. Similar to the previous example, the committee includes a set of instructions to constrain the activities of the Commission to specific areas. Although the committee commends the Commission on its work in these area, it encourages the agency to continue its efforts in specific activities – thus effectively constraining its ability to do other things.

> The number of issues affecting marine mammal conservation has grown over the past three decades and the Committee recognizes these increasing responsibilities of the Commission As such, the Committee recommendation includes $3,200,000 for the necessary expenses, including the hiring of two additional FTEs, for the Marine Mammal Commission. This amount is $380,000 above the current year level and $800,000

> above the request. The Committee has provided this increase to allow the Commission to address various challenges including the effects of and adaptation to climate change and the study of endangered species. The additional FTEs would be responsible for monitoring oil and gas issues such as the expansion of exploration and clean-up of oil spills on ice as well as emerging fishery issues.
>
> The Committee understands that marine mammals in polar regions are particularly vulnerable to the effects of climate change due to expected reduction in sea ice and subsequent changes in ecosystem biodiversity and structure. Many of these animals are poorly studied and already may be experiencing serious effects from climate change. The Commission has initiated a project to develop monitoring strategies for Arctic marine mammals and the Committee supports efforts to complete this work and then to shift focus to development of mitigation efforts.
>
> The Committee expects the Commission to continue its efforts to minimize the direct and indirect effects of fisheries, noise, disease, chemical contaminants, harmful algal blooms, climate change, habitat alteration, boating and commercial shipping, marine debris, and other factors that may pose a risk of sublethal and lethal effects on marine mammals or that may affect the health and stability of the marine ecosystem.

On the other hand, instructions to the Marine Mammal Commission in House Report 108-576 in FY 2005 is substantially less constraining. In fact, the committee only uses one sentence to recommend funds to the agency without any instruction:

> The recommendation includes $1,890,000 for the necessary expenses of the Marine Mammal Commission, which is $54,000 above the fiscal year 2004 level and the same as the request.

Overall, this pair of examples demonstrates a number of features concerning these committee reports. First, it shows situations where reports can be more or less constraining. Second, it highlights the variation that can occur over time. Even within a single agency, a subcommittee can still choose to change the level of specificity. This also provides evidence that committee reports are not completely path dependent, simply repeating language from past years. In other words, decisions on constraints within reports are based upon a number of other political factors and not just on the content of the previous report. Finally, the example demonstrates how a shift from unified to divided government can influence these decisions. In particular, the examples show a more constraining report made under divided government (FY 2009) and a less constraining one under unified (FY 2005).

B Reliability of CQ Summaries

In this appendix, we review the issue of the reliability of CQ summaries in greater detail. These summaries serve as the backbone for the Epstein and O'Halloran measure of discretion as well as a number of other measures in different studies. The summaries, while usefully condensing general information about legislation, may also be incomplete in ways that are problematic for measures of discretion.

To examine this question, we focus on two bills – one that the EO measure suggests had zero delegations and one that the measure suggests had zero constraints. We focus on these cases because they most clearly demonstrate the issues we raise here. A closer examination of the statutes suggest that the characterizations of no delegation or no constraint in the summaries may not necessarily match the statutory language.

Consider, for example, the Revenue and Expenditure Control Act of 1968 (PL 90-364). The EO measure records this bill as having zero delegations, and thus zero discretion. One key provision of this act was to cap the total number of federal civilian employees at the level on June 30, 1966. This employee cap and others like it throughout the 1950s, 60s, and 70s had a significant impact on how the government was able to carry out its functions (see Light 1999, for an extended discussion). Perhaps most significantly for our purposes, the legislation delegates substantial authority to the Bureau of the Budget to determine what agencies should gain and lose employees under the cap, and gave the Director of BOB wide latitude in making that determination:

> For the purposes of paragraph (1) [which lays out the numerical limit on employees], the Director may reassign vacancies from one department or agency to another department or agency when such reassignment is, in the opinion of the Director, necessary or appropriate because of the creation of a new department or agency, because of a change in functions, or for the more efficient operation of the government.

This relatively vague grant of authority is consequential due to the importance of personnel for the capacity and production of agencies (see, for example, Carpenter et al. 2012; Bolton, Potter and Thrower 2016) as well as evidence that suggests presidents act ideologically in their allocation

of personnel resources across agencies. Furthermore, the act gives the Director of BOB significant regulatory authority to carry out the employment cap policy ("The Director shall prescribe such regulations as he deems necessary or appropriate to carry out the provisions of this section"). The employment cap provision is not the only instance of what we would argue count as delegations in this legislation. For instance, there are provisions in Title I that grant the Secretary of the Treasury power to exempt corporations from some tax requirements in the bill at his or her discretion (e.g. Section 6425(d)). These relatively clear delegations of authority are missing from the CQ summaries.

Now consider a bill that, by the EO measure has a Discretion Index of one, indicating delegation in all provisions and no constraints – the 1961 Inter-American Program Appropriation (PL 87-41). The purpose of this bill was to appropriate funds for the newly created Inter-American Program (which eventually became the Inter-American Foundation). The program was established to fund development projects in Latin American countries. We argue that this appropriation bill does, in fact, include constraints on how the program could use the money. The text of the statute itself includes a constraint directly after the legislation appropriates the money [emphasis added]:

> For expenses necessary to carry out the provisions of sections 1 and 2 of the Act of September 8, 1960 (74 Stat. 869), $500,000,000, to remain available until expended: ***Provided*, That the funds herein appropriated shall not be available to be loaned or reloaned at interest rates considered to be excessive by the Inter-American Development Bank or higher than the legal rate of interest of the country in which the loan is made.**

This has a significant impact on the discretionary zone in which the program could operate, though it would not be counted as a constraint in the EO measure because it is not focused on procedures. However, there are procedural (and other) constraints placed on the use of the appropriation in the Senate Appropriation Committee's report on the legislation. Indeed, the report (12322 Senate Report 201) backs up the statutory language on interest rates and other constraints with the imposition of a reporting requirement on the program administrators, which is a constraint recognized by the EO measure:

> The Committees on Appropriations of the Senate and the House of Representatives should be provided with semiannual reports on the loans made with the funds appropriated in this bill. If exceptional cases should develop in connection with loans where it is necessary to limit first-year payments to interest such exceptional cases should be reported in the aforementioned semiannual report, together with the reasons why exceptions were made to the policy outlined herein.

These examples raise questions about the completeness of the CQ summaries. It is unclear whether or not these omissions are systematic, raising further questions about how they might impact measures of discretion and the types of inferences that we can make from them. While the summaries are a convenient and easily available method of coding legislation, they may or may not capture the relevant aspects of the concept of discretion that would be necessary to create a reliable measure of discretion.

C Accounting for Presidential Budget Requests

One concern that may arise in the use of the proposed measure of discretion is that Congress may be responsive to presidential requests, particularly with respect to the page length of reports. Schick (2008) argues that when presidents propose particularly large changes to appropriations, Congress potentially writes longer reports.[1] If this were the case, it may artificially increase the level of constraint that we measure, thus reducing the amount of measured discretion that Congress is giving to an agency.

In order to account for this possibility, we reconsider our ally principle analysis, taking into account the change in appropriations being requested by presidents. In particular, we control for the logged absolute value of the president's proposed percentage appropriations change from the current fiscal year. As this value increases, the president is seeking larger changes to appropriations, and potentially leading to longer appropriations committee reports. One complication that arises in the analysis of the appropriations process is that presidents may be strategic in their requests, anticipating the responses of Congress and adjusting accordingly in order to get the appropriation closest to their ideal point. In the existing literature (e.g. Howell and Jackman 2011; Kiewiet and McCubbins 1991), scholars have sought to instrument for the magnitude of the president's request using indicators for the president's first term as well as for the first, second, and third years of the president's term. We also pursue this strategy.

Table 1 below replicates Model 4 in Table 1 in the manuscript. Model 1 includes the absolute magnitude of the president's requested change in the ally principle regression. Model 2 includes the results from an analysis where we instrument for this value using the year and term instruments from the previous literature on appropriations. The instruments are strong, with a Cragg-Donald Wald F-statistic of 20.5. Overall, the results suggest that the ally principle still is evident in the data, even after accounting for the size of a president's request.

In Model 1, we continue to see a negative relationship between ideological divergence and

[1]This assertion is not tested empirically in Schick (2008).

Table 1: Ally Principle Test Accounting for Presidential Requests

Variable	Model 1	Model 2
Chamber-President Disagreement	-0.130***	-0.174***
	(0.037)	(0.044)
Presidential Request	-0.028	0.090
	(0.017)	(0.065)
Inflation	0.016	0.019
	(0.013)	(0.013)
Unemployment	-0.006	-0.005
	(0.014)	(0.014)
War	-0.063	-0.053
	(0.048)	(0.049)
Public Approval	0.004**	0.004***
	(0.002)	(0.002)
Agency Fixed Effects	✓	✓
Chamber-Subcommittee Fixed Effects	✓	✓
N	8,112	8,112

Coefficients reported from OLS regression model, with robust standard errors clustered by agency in parentheses. Model 2 includes instrumented values of presidential requests. Significance codes: * p < 0.10, **p < 0.05, ***p < 0.01, two-tailed tests.

discretion, where chamber disagreement is associated with a 12.2% decrease in discretion levels. There does appear to be a negative correlation between the magnitude of the president's request and discretion, as would be predicted by Schick (2008). Moving to the instrumental variables analysis reported in Model 2, we still see the negative effect of ideological disagreement, with an estimated 16% decrease in discretion. This estimate is both statistically significant and substantively in line with the estimates reported throughout the paper. Here the effect of presidential request magnitude actually appears to be positive, however, the standard error on the estimate is quite large. Overall, then, these results suggest that the ally principle still holds, even after accounting for presidential activity.

Another way in which strategic considerations may enter decisions is when the House and the Senate are controlled by different parties. There are a number of ways that this ideological con-

Table 2: Ally Principle Test Accounting for Divided Chambers

Variable	Model 1
Chamber-President Disagreement	-0.154***
	(0.041)
Divided Chambers	-0.079*
	(0.046)
Inflation	0.019
	(0.013)
Unemployment	0.003
	(0.014)
Administration Change	-0.063**
	(0.030)
War	-0.079*
	(0.048)
Public Approval	0.005***
	(0.002)
Agency Fixed Effects	✓
Chamber-Subcommittee Fixed Effects	✓
N	8,284

Coefficients reported from OLS regression model, with robust standard errors clustered by agency in parentheses. Model 2 includes instrumented values of presidential requests. Significance codes: * p < 0.10, **p < 0.05, ***p < 0.01, two-tailed tests.

figuration could affect observed outcomes. For example, the chambers may write systematically longer reports in order to ensure that agencies comply with their instructions. They might also propose more or less appropriations in anticipation of a battle with the other chamber. To account for this possibility, we replicate the analysis of the ally principle while adding an additional control for times in which there different parties control each chamber. The estimated coefficients for this analysis are reported in Table 2.

As can be seen, even after controlling for ideological divergence between the House and the Senate, the effect of Chamber-President Disagreement remains substantively identical. Moving from ideological agreement to disagreement with the president results in a 14.3% decrease in discretion. Interestingly, the divided chamber variable is negative and marginally significant (p =

0.089), suggesting that discretion is decreased during these periods of disagreement, perhaps as chambers attempt to stymie the influence of the other on executive branch agencies.

D Reliability of Measure Across Agencies and Time

In this section, we review a number of robustness checks that examine whether our proposed measure of discretion is relatively consistent across agency types and time. While we use agency fixed effects in all of our models throughout the paper (thus identifying effects within agencies), it is important to show that no particular agency or group of agencies behaves significantly differently from others with respect to our measure. This is important for establishing the ability to use this measure in analyses with both many agencies as well as particular subsets of agencies.

Our approach here is to run a series of analyses that replicate Model 4 of Table 1 in the main text of the paper and examine the stability of the results from our ally principle analysis accounting for time and agency differences in a variety of ways. Overall, the results we report below are remarkably stable, suggesting that the patterns we show as evidence for the construct validity of our measure are not driven by any one agency, set of agencies, or time period. This demonstrates the utility of the measure for potential users in a variety of applications. In sum, we conduct the following analyses:

- Separate out agencies that have alternative funding structures that could impact the applicability of our measure, including the ability to collect fees and other receipts and use these funds to finance substantial portions of agency activities as well as the ability to draw on working capital funds

- Iteratively drop agencies in major policy areas to confirm that effects are not driven by particular types of activities

- Isolate agencies that engage in substantial amounts of economically significant regulation and those that do not to confirm that effects are not driven by this activity

- Subset agencies by size to ensure that the results hold up equally in different groups

- Include agency-specific time trends with varying degrees of polynomials (linear, quadratic, and cubic) to demonstrate that over-time results within agency are stable

- Examine dynamic effects of the measure
- Standardize the measure within agencies

To begin, we examine whether the ally principle results are stable across a variety of agency funding structures. As previously noted, some agencies do not rely wholly on annual appropriations for their operations but instead draw upon proceeds from user fees, working capital funds, or other forms of receipts in order to fund their operations or parts thereof. To be sure, even agencies that are funded by these alternative sources are most often bound by obligation limitations imposed through the appropriations process (which we record as the level of delegation in our measure). Given that our measure is focused on the appropriations process, however, we want to take special care to ensure that the measure and results are consistent across these types of arrangements.

In particular, we focused on three characteristics of agency funding structures salient for our measure: (1) whether or not the agency is able to assess and collect reasonable fees for its activities; (2) whether or not the agency and its activities are funded at least in part by a working capital fund giving the agency access to the money irrespective of other limitations on spending; and (3) whether or not an agency or part of an agency uses assessed fees or other receipts in order to substantially fund its total operations. We identified agencies with these funding features using the Administrative Conference of the United States's *Sourcebook of United States Executive Agencies* (Lewis and Selin 2012).

All three of these features may, in theory, substantially alter the meaning of appropriated funds relative to other agencies where these conditions do not apply. In order to assess whether the key ally principle result nonetheless holds under these conditions, we subsetted our dataset into agencies that have these features and those that do not and re-ran Model 4 in Table 1 on each sample. The estimated coefficients from these analyses are reported in Table 3. Across these six models, we see strong support for the ally principle. There is a consistent, negative, and statistically significant impact of ideological divergence between the president and the chamber on the levels of discretion granted to agencies. Importantly, the magnitude of the estimated effects is relatively consistent with those reported in the main text across all six models. The lone exception is among

Table 3: Ally Principle Robustness – Various Appropriations Structures

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Chamber-President Disagreement	-0.149***	-0.176**	-0.158***	-0.159	-0.107***	-0.320***
	(0.0438)	(0.0645)	(0.0424)	(0.107)	(0.0393)	(0.0772)
Inflation	0.0161	0.0256	0.0129	0.0570	0.00766	0.0585*
	(0.0147)	(0.0187)	(0.0131)	(0.0399)	(0.0119)	(0.0335)
Unemployment	-0.0181	0.0436*	-0.0140	0.0692	-0.0107	0.0238
	(0.0169)	(0.0230)	(0.0148)	(0.0471)	(0.0163)	(0.0250)
Administration Change	-0.0672**	0.0410	-0.0589*	0.0414	-0.0403	-0.0792
	(0.0322)	(0.0666)	(0.0319)	(0.101)	(0.0327)	(0.0624)
War	-0.0551	-0.123*	-0.0730	-0.0222	-0.0234	-0.234**
	(0.0536)	(0.0707)	(0.0499)	(0.174)	(0.0457)	(0.102)
Public Approval	0.00491***	0.00370	0.00479***	0.00131	0.00466***	0.00434
	(0.00174)	(0.00336)	(0.00167)	(0.00526)	(0.00169)	(0.00427)
Agency Sample	No Fees	Fees	No Offsets	Offsets	No WCF	WCF
Agency FE	✓	✓	✓	✓	✓	✓
Chamber-Subcommittee FE	✓	✓	✓	✓	✓	✓
N	6711	1573	7566	718	6864	1420

Coefficients reported from OLS regression model, with robust standard errors clustered by agency in parentheses. Significance codes: *p < 0.10, **p < 0.05, ***p < 0.01, two-tailed tests.

the group of agencies that use fees to offset substantial portions of their expenses, where we find only a marginally significant effect ($p = 0.16$). However, it should be noted that this is a relatively small group of agencies (twelve) and the effect is in the expected direction. The magnitude of the effect is also fairly large, alleviating concerns that this group of agencies is fundamentally different in some way with respect to our discretion measure. In all, these results help to lend credence to the idea that agency funding structures are not affecting the applicability of our measure.

Next, we turn to the issue of whether or not particular activities that agencies are engaged in substantially alter the applicability of the measure. It is possible that agencies in some policy areas or engaged in particular types of activities may use money more or less efficiently, leading to concerns that the delegation portion of the measure has different meanings across organizations and therefore will perform differently. First, we examined whether or not particular policy areas were driving the results that we reported in the main text of the paper. To do this, we first assigned one of twenty-one major policy areas (as coded by the Policy Agendas Project) to each agency based on their mission statement. Then, we re-ran Model 4 from Table 1 in the main text of the paper, while iteratively dropping agencies in each policy area. The goal of the analysis is to assess whether the results are stable when dropping each set of agencies. If they are, this suggests that our measure is capturing similar dynamics across policy areas.

Figure 1 displays the estimated coefficients and 95% confidence intervals for the *Chamber-President Disagreement* variable with agencies primarily working in the policy area indicated on the x-axis omitted. The most notable and immediately visible pattern is that the estimated coefficients are extremely stable across all of the samples. No matter which policy area is dropped, the estimated effect is negative and statistically significant. Furthermore, the magnitude of the effect size is extremely stable across all of the policy areas. The biggest discrepancies appear when dropping the domestic commerce and government operations policy areas (since the effect decreases in magnitude the most when those categories are omitted). Nonetheless, even when they are omitted, the statistically and substantively significant effect persists. These results provide evidence that suggests that the main policy area of activity for an agency does not appear to impact the

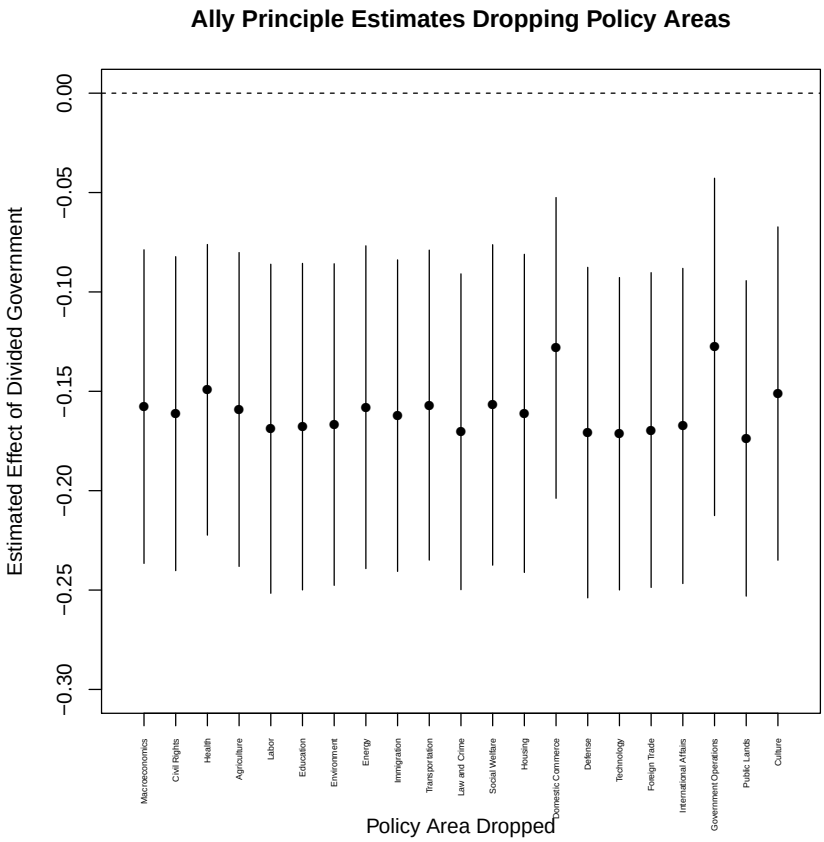

Figure 1: Estimated coefficients and 95% confidence intervals for the *Chamber-President Disagreement* variable while iteratively dropping agencies working in particular policy areas.

performance of our measure.

In addition to policy areas, we also attempt to determine whether the functions that agencies carry out may lead to different results with our measure. In particular, we focus on whether or not agencies engage in economically significant rulemaking on at least a semi-regular basis. Rulemaking is of particular interest because it can lead to intense policy conflict between the executive branch and Congress. One could imagine that this type of policymaking activity might be particularly applicable for our measure (given the role of limitation riders to appropriation bills in constraining regulations) while other agencies may not quite fit with the appropriations discretion formulation given their lack of a significant regulatory function.

To examine this possibility, we use Lewis and Selin (2012) to identify agencies that engaged

in economically significant rulemaking during the sample period of that study. We re-ran Model 4 from Table 1 on agencies that engage in this type of rulemaking and those that do not in order to assess whether the effect of ideological disagreement is consistent across the sample. Table 4 displays the results from this analysis. As seen, the ally principle holds for both sets of agencies. Furthermore, the differences in effect sizes are negligible, again suggesting that our measure does not perform differently in agencies that differ in terms of the degree to which they engage in regulation.

Table 4: Ally Principle Robustness – Agencies Issuing Economically Significant Regulations

	Model 1	Model 2
Chamber-President Disagreement	-0.152***	-0.171**
	(0.0450)	(0.0606)
Inflation	0.0233	0.00751
	(0.0154)	(0.0204)
Unemployment	-0.0164	0.0306
	(0.0180)	(0.0189)
Administration Change	-0.0630**	0.00572
	(0.0312)	(0.0611)
War	-0.0781	-0.0457
	(0.0574)	(0.0755)
Public Approval	0.00494**	0.00254
	(0.00198)	(0.00272)
Issue Econ. Sig. Regs	No	Yes
Agency FE	✓	✓
Chamber-Subcommittee FE	✓	✓
N	6322	1962

Coefficients reported from OLS regression model, with robust standard errors clustered by agency in parentheses. Significance codes: *p < 0.10, **p < 0.05, ***p < 0.01, two-tailed tests.

Finally, we also examine whether agency size impacts the ways in which the measure works. To do this, we divided agencies into "small" and "large" in two different ways. First, we used information about the size of the agencies budget to classify agencies as small (on average less than $100 million) and large (on average greater than or equal to $100 million). The second way

we examined this was to separate out agencies that were Cabinet departments from those that were not. The Cabinet departments are among the largest organizations both in terms of money, personnel, and the importance of their missions – making this is an appropriate distinction. A final way in which we seek to assess scale effects is to control for the logged level of budget authority in the current fiscal year.

As in the other analyses, we replicate the results for Model 4 in Table 1 to evaluate the degree to which agency size impacts the results. Table 5 displays the results of these analyses. The results clearly track those in the main paper, in line with the ally principle no matter which of the four samples we examine. The findings based on the size of the agency budget (reported in Models 1 and 2) are nearly identical and consistent with the ally principal. With respect to the Cabinet/non-Cabinet samples (results reported in Models 3 and 4), both yield estimates that are negative and statistically significant, consistent with the ally principle. The magnitude of the effect size is larger for Cabinet agencies (-0.28) than non-Cabinet agencies (-0.12), perhaps reflecting the importance and salience of their missions relative to agencies outside the Cabinet. In Model 5, we use all observations to observe the degree to which current appropriations impact the stability of our results. As can be seen, we estimate a positive effect of current year appropriations. Importantly, the effect of divided government remains precisely estimated, albeit somewhat smaller in magnitude than the results reported in the main text. In sum, the relative stability of the effects across these different groups of agencies and controlling for within agency changes in size provide further evidence of the consistent performance of the measure in a variety of organizational contexts.

The next concern that we address pertains to the degree to which the measure produces consistent, reliable results over time within an agency. In order to assess the degree to which the measure is sensitive to these concerns, we account for time in a variety of ways for Model 4 in Table 1 reported in the main text. The results of these analyses are reported below in Table 6. First, and most simply, we include a global, linear time trend to account for any potential trends in levels of discretion. For instance, one might imagine that other trended variables, such as polarization, might impact overall levels of discretion. Model 1 in Table 6 reports the estimated coefficients from

Table 5: Ally Principle Robustness – Agency Size

	Model 1	Model 2	Model 3	Model 4	Model 5
Chamber-President Disagreement	-0.166***	-0.146**	-0.116***	-0.285***	-0.0935***
	(0.0484)	(0.0623)	(0.0405)	(0.0845)	(0.0264)
Inflation	0.00503	0.0293	0.0159	0.0184	0.0138*
	(0.0160)	(0.0194)	(0.0133)	(0.0298)	(0.00798)
Unemployment	-0.0233	0.0173	-0.0117	0.0260	0.00456
	(0.0210)	(0.0182)	(0.0178)	(0.0190)	(0.0109)
Administration Change	-0.0804**	-0.00694	-0.0541*	-0.00300	-0.0580**
	(0.0347)	(0.0463)	(0.0321)	(0.0563)	(0.0245)
War	-0.0253	-0.103	-0.0373	-0.135	-0.113***
	(0.0593)	(0.0707)	(0.0494)	(0.105)	(0.0378)
Public Approval	0.00467**	0.00355*	0.00412**	0.00338	0.00585***
	(0.00226)	(0.00207)	(0.00194)	(0.00269)	(0.00129)
Ln(Current FY Appropriations)					0.473***
					(0.0401)
Agency Sample	Small Budget	Large Budget	Non-Cabinet	Cabinet	All
Agency FE	✓	✓	✓	✓	
Chamber-Subcommittee FE	✓	✓	✓	✓	
N	4685	3599	6469	1815	8112

Coefficients reported from OLS regression model, with robust standard errors clustered by agency in parentheses. Significance codes: *p < 0.10, **p < 0.05, ***p < 0.01, two-tailed tests.

this regression, showing that the global time trend coefficient is small and statistically insignificant. Moreover, the estimated impact of chamber disagreement remains negative and statistically significant, at a magnitude comparable to that when the trend variable is omitted.

However, one might be concerned that a global time trend is not the appropriate way to measure this relationship. In particular, time trends might differ significantly depending on the agency that we are examining. For instance, some agencies in our dataset, like the Interstate Commerce Commission, slowly declined and were ultimately terminated, while others have had the opposite fate. To allow for this flexibility in accounting for time, we add agency-specific time trends to Model 4 in Table 1. In order to assess the robustness of this approach, we include these trends linearly as well as with quadratic and cubic specifications to allow for flexibility in the impact of time; we report the results in Models 2-4 in Table 6, respectively. No matter what the functional form of the time trend, we continue to find a statistically and substantively significant and negative estimated coefficient for the ideological divergence variable consistent with the ally principle.

We conduct one final analysis in order to assess the impact of the issuance of the Supreme Court decision in *INS v. Chadha*, which ruled that legislative vetoes were unconstitutional violations of the presentment clause. As Fisher (2005) reports, legislative vetoes have not actually gone away since the Supreme Court case. Instead, many moved into appropriations committee reports where their constitutionality could not be challenged since they were not technically law. If this is true, then we would expect that post-*Chadha* appropriations reports might increase in length due to this change and decrease discretion across the board. To investigate this possibility, we include an indicator for fiscal years beginning in 1984, *Post-Chadha*. These results are reported in Model 5. As can be seen, there does seem to be a decrease in discretion post-Chadha; however, the ally principle still manifests with the inclusion of this variable. This helps to provide assurance that even when taking into account time in a number of flexible ways (i.e. global trends, polynomials and specific to agency, and potential structural breaks), the measure performs reliably, producing similar results across specifications.

One final issue related to time is the degree to which there are dynamic effects in the bud-

Table 6: Ally Principle Robustness – Reliability Over Time

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Chamber-President Disagreement	-0.148***	-0.156***	-0.116***	-0.107***	-0.107***	-0.0569***	-0.0684***
	(0.0392)	(0.0349)	(0.0302)	(0.0289)	(0.0350)	(0.0213)	(0.0207)
Inflation	0.0345***	0.0247**	0.0121	0.0109	-0.0264**	0.0102*	0.0170**
	(0.0102)	(0.00962)	(0.00843)	(0.00850)	(0.0121)	(0.00592)	(0.00765)
Unemployment	0.00522	-0.00511	-0.00390	-0.000278	-0.0105	-0.00968	0.000288
	(0.0122)	(0.0107)	(0.0107)	(0.0107)	(0.0145)	(0.00745)	(0.00878)
Administration Change	-0.0508*	-0.0754***	-0.0477*	-0.0486*	-0.0517*	-0.0175	-0.00919
	(0.0302)	(0.0278)	(0.0275)	(0.0279)	(0.0298)	(0.0269)	(0.0214)
War	-0.112**	-0.0946**	-0.0416	-0.0253	0.0269	-0.000489	-0.0500*
	(0.0440)	(0.0431)	(0.0392)	(0.0394)	(0.0447)	(0.0268)	(0.0287)
Public Approval	0.00563***	0.00620***	0.00459***	0.00416***	0.00307**	0.00314***	0.00319***
	(0.00148)	(0.00162)	(0.00144)	(0.00140)	(0.00152)	(0.00108)	(0.00110)
Trend							
Post-Chadha					-0.443***		
					(0.106)		
Lagged Discretion						0.597***	0.132***
						(0.0461)	(0.0391)
Agency-Specific Trend Functional Form	N/A	Linear	Quadratic	Cubic	N/A	N/A	N/A
Agency FE	✓	✓	✓	✓	✓	✓	
Chamber-Subcommittee FE	✓	✓	✓	✓	✓		
N	8284	8284	8284	8284	8284	6823	5664

Coefficients reported from OLS regression model, with robust standard errors clustered by agency in parentheses. Significance codes: *p < 0.10, **p < 0.05, ***p < 0.01, two-tailed tests.

geting process that need to be accounted for and whether the ally principle results in the paper hold up to this complication. In particular, the budget process is often described as incremental (Davis, Dempster and Wildavsky 1966).[2] Thus we might expect that budget authority in the previous period may impact it in the next period. Such stickiness may not be as clearly applicable to the constraint portion of the measure. In order to examine this, in Models 6 and 7 in Table 6 we estimate models that account for the dynamics in discretion. Model 6, includes a lagged dependent variable, which allows us to estimate the long- and short-run effects of divided government. The estimated coefficient on the lagged discretion variable is approximately 0.597, implying an adjustment coefficient of 0.403. However, Model 6, in including both a lagged dependent variable and fixed effects, is potentially susceptible to Nickell bias for small T panels. To account for this possibility, we also use the Arellano-Bond estimator in Model 7 of Table 6 and estimate similar effects.[3]

Finally, if one is still wary of potential inconsistencies in the measure across agencies (and that variation is not of interest in their particular application), one can also standardize the measure within agencies, essentially creating within agency z-scores. This allows the user to purge the differences in mean levels of discretion and account for the variation within agencies before carrying out their analysis. If one does this, the results of the ally principle still hold. Estimating Model 4 in Table 1 with this standardized measure of discretion yields an estimated coefficient of -0.244 (standard error = 0.038) for the chamber disagreement variable. Substantively, this effect corresponds to a 0.244 standard deviation decrease in appropriations discretion when the chamber is ideologically-distant from the president relative to when they are controlled by the same party. Alternatively, demeaned versions of the measure could also be employed to purge an agency-, committee-, or time-specific variation in the measure. This would likely be particularly attractive for individuals using the discretion measure as an independent variable.

[2]See, however, Padgett (1980) and Jones, Zalányi and Érdi (2014) for alternative views.

[3]Note that the fixed effects are differenced out in this specification.

In all, the results reported in this section alleviate concerns about the reliability of our proposed measure and its comparability across different types of agencies and time. Our approach has been to show that our main test of the construct validity of our measure – the ally principle – holds up in a variety of contexts, using different subsets of agencies and accounting for time in a number of ways. The tests where we subset our dataset on various agency characteristics – funding structures, policy area, rulemaking activities, and agency size – show the consistent performance of the measure across a wide range of samples of agencies chosen based on politically salient characteristics. The results were consistent not just in terms of sign and statistical significance but in terms of the magnitude of the coefficients as well. Additionally, we included time in the analysis in a variety of ways that were flexible, both in in the sense of varying across agencies and in functional form. The consistent support for the ally principle across these analyses further supports the validity of our measure.

E Accounting for Authorizations

One concern that may arise from the use of our measure is that it is based on appropriations and neglects other parts of the legislative process that can also affect an agency's discretion. In the main text of the paper, we discuss a number of reasons why appropriations is a major venue in which agency discretion is determined in an ongoing way. Additionally, a measure based on appropriations is ideal for testing theories of discretion because of its salience for agency functioning, its repeated nature that helps to eliminate selection problems, and the relatively clear nature of appropriations decisions. However, most agencies are created and also operate under conditions set by their authorizing legislation. In this section, we consider the degree to which our measure is affected by activity in the authorizing process.

There are a number of ways in which the authorizations process itself might affect our measure. First, it is possible that discretion in the authorizations and appropriations processes are substitutes because legislators who constrain an agency in the authorization process may not need to further constrain it in the appropriations process, and vice versa. In that case, we would expect authorizations discretion and appropriations discretion to be negatively correlated with one another. If this were the case, however, we would be unlikely to observe the ally principle, which has been demonstrated in the authorizations-based measures of discretion. Observing similar key results suggests that the measures may themselves be positively correlated, providing evidence they reflect similar political dynamics.

There are a number of ways through which the authorizations process could still impact our measure. It is possible that the reports written by the appropriations committees often replicate some of the instructions of the authorizing legislation. However, the frequently documented policy conflicts between the appropriations and authorizing committees appears to belie this particular possibility to some degree. Furthermore, we would argue that emphasizing particular aspects of authorizing legislation – such as programs or procedures – may actually be constraints that we would want our measure to capture. Here, the committee is urging the agency to prioritize certain activities. Given the limited financial, attention, and other resources that agencies possess, this nec-

essarily means deemphasizing other activities. Furthermore, new delegations from the authorizing process would likely show up in our measure as increased appropriations.

A weaker version of this possibility is that when the appropriations committees write reports, they are responsive to activity in the authorizations process for an agency. If this is the case, then we would expect a strong, positive relationship between the length of reports and authorizing activity in an agency's jurisdiction. In order to investigate this possibility, we attempt to characterize authorizing activity in an agency's area in three ways, by collecting information on the number of all laws that have been passed in an agency's policy jurisdiction, the number of "significant" laws (as defined by Mayhew (2005)) that have been passed for which the agency was the primary recipient of delegated authority recently, as well as the number of expired authorizations for which there are appropriations in an agency's jurisdiction in a given fiscal year.

To create the first measure, we used the Policy Agendas Project's (PAP) twenty one major policy codes to characterize the policy areas in which an agency works. Note that agencies can have multiple codes assigned to them based on their mission. For instance, the Department of Agriculture is assigned to the "agriculture" policy area but also to the "health" policy area for its work on school lunch programs, the "social welfare" and "housing" codes due to its work on rural housing and other aid to rural communities, the "foreign trade" policy area due to its work through the Foreign Agricultural Service, and to the "public lands" category for its jurisdiction over all federal forests via the Forest Service.[4] Most agencies have only one policy area assigned to them, but many of the Cabinet departments are active in a variety of policy areas, which is what we hope to capture by coding agencies in this way. We then use the data from PAP that categorizes public laws by major policy code to create a measure of the number of public laws passed in an agency's jurisdiction in the previous year, the previous three years, and the previous five years. We refer to

[4]Note that we also assigned agencies to one "primary" area to assess the robustness of the results we report below to a more restrictive coding. The results and discussion below are substantively the same whether we using the inclusive or restrictive coding scheme for this measure.

these measures as *Laws Passed*, *Laws Passed, 3 Yr.*, and *Laws Passed, 5 Yr.* below.

However, many laws passed by Congress are relatively minor or limited in scope. Because of this, using counts of any laws passed in an agency's jurisdiction as a measure of new policy authorities may not fully capture the concept of interest. To remedy this, we also focus on a subset of legislation that has been deemed "significant" by Mayhew in his work *Divided We Govern* and in subsequent updates of the measure. We examined each law to determine the agency that was primarily responsible for administering it. In some cases, there were a number of agencies that were granted authorities under the law, but for this analysis we isolate the agency that received the primary bulk of the delegations and responsibility. For example, the 2010 Patient Protection and Affordable Care Act (i.e. "Obamacare") primarily created programs that were to be administered by the Department of Health and Human Services; however the law also contained new authorities for the Architectural and Transportation Barriers Compliance Board, Department of Defense, Department of Labor, Department of the Treasury, Department of Veterans Affairs, the Independent Payment Advisory Board, Medicaid and CHIP Payment and Access Commission, Office of Personnel Management, Social Security Administration, and US Sentencing Commission. Given the responsibilities for these other agencies were far less expansive than for HHS, we code HHS as the primary agency for the law.[5] In a small number of cases, we also recorded more than one primary agency. For example, the Omnibus Public Land Management Act of 2009 deals significantly both with national forests (administered by the Forest Service in the Department of Agriculture) as well as other public lands falling under the purview of the Bureau of Land Management in the Department of the Interior. Here, we coded both agencies as primary for the purposes of our analyses.

The number of laws in an agency's policy area that are passed exhibits great variation over time and within agency. The modal number of laws passed in an agency's policy area is zero, but some

[5]It is possible to conduct a similar analysis using all of the agencies implicated by Mayhew's laws and the results we report here are not substantively altered.

agencies with many policy areas in their jurisdiction, such as Cabinet departments or over-arching policy agencies like the Office of Policy Development in the Executive Office of the President can have over one hundred in a given year. In order to account for this huge skew, we use the natural logarithm of these variables in the analyses reported below. Unsurprisingly, the numbers using only significant legislation are much smaller, with the maximum number for an agency in a year being 5. The mode, however, is still zero and the distribution skews substantially toward zero, leading us to use a natural logarithmic transformation for this data as well.

The final measure that we create to characterize authorizations activity is the number of programs in an agency's jurisdiction in a given year for which authorizations have expired and budget authority has been appropriated. This information was collected from the Congressional Budget Office's annual report to Congress, typically titled "Unauthorized Appropriations and Expiring Authorizations." This annual report, mandated by amendments to the Budget and Accounting Act of 1974, is available from fiscal year 1988 onward.[6] In this report, CBO lists (by committee) the programs that have expired authorizations. We matched each program to the agency that oversees it to generate a count of unauthorized programs for each agency-year. The modal number of unauthorized programs is zero, but goes as high as 142 for some agencies and years. Given the skew in the variable, we use the natural logarithm of it in our analyses.

To begin exploring the relationship between our discretion measure and these measures of activity in the authorizations process, we first examine the correlations between each measure and the number of pages devoted to agencies in the appropriations committee reports. Given that we use the natural logarithm of the authorizations measures, we also use the natural logarithm of pages when reporting these correlations, which also reflects their form in the discretion formulation that we ultimately adopt.

---

[6]The report was first published for FY1986 but did not include precise amounts of unauthorized funds until FY 1988. In order to be sure that all programs and amounts are included, we begin in FY 1988 when this information is wholly inclusive.

When examining the correlations between the appropriations process, as measured by the committee reports, and our three measures of the authorization process, one could reasonably expect these processes to be either complements or substitutes based on different expectations about legislators' behavior towards agencies. On the one hand, legislators could use the committee reports to appropriations bills to give further instruction to agencies already receiving attention in the authorization process. If that were the case, then we would expect more legislation devoted to an agency (and fewer expired authorizations) to be associated with *longer* committee reports. On the other hand, legislators could use attention to an agency in the appropriations process as a substitute for the lack of attention given to it in authorization process, and vice versa. As such, we would expect more legislation to an agency (and fewer expired authorizations) to be associated with *shorter* committee reports.

To explore these conflicting expectations, we calculate the correlations between the logged number of pieces of legislation in the agency's jurisdiction and the logged number of pages devoted to the agency in reports, which is 0.19, suggesting that when there is more authorizations activity, appropriations committees are responsive to this in the reports in at least a limited way, writing longer reports. This complementary relationship is mirrored in the correlation between the number of significant laws where the agency is the primary one and page numbers, which is 0.27. The correlation between the logged number of expired programs and page length, 0.67, appears to contradict this at first blush. This large, and strong correlation suggests that when agencies have expired authorizations, the appropriations committee fills the gap as a substitute in the absence of legislation. All three of these correlations are significant at the $p < 0.001$ level. While one would expect these two types of measures of authorizations activity (law production and expirations) to be negatively correlated, the correlation between laws in an agency's policy area and expiring provisions is 0.15. How can these results be squared?

In general, we argue that it is important to take into account agency differences when considering these results. Large agencies have the most programs and longest reports as well as the largest numbers of expired authorizations and laws in their jurisdictions. If we do not take this

into account, these correlations are highly misleading. Therefore, we demean the length of the reports, the number of authorizing laws in the agency's policy areas, and the number of programs with authorized programs in the agency's jurisdiction by agency and reconsider the correlations between these factors. Now, a more coherent picture emerges. The correlation between pages and the number of laws in an agency's jurisdiction is small and negative, -0.040, suggesting that there is little correspondence with authorizations activity and appropriations. If anything, they appear to be somewhat substitutable but to a very small degree. Similarly, the correlation between the Mayhew-based measure and pages after demeaning also shrinks considerably to 0.03. A somewhat more robust correlation appears to exist between demeaned number of unauthorized programs and report lengths, 0.12. This suggests when the authorizations process breaks down, the appropriations committee appears to step into the breach to a modest degree.

The relatively small correlations between page lengths and authorizations activity provides some reassurance that activity in the authorizations process is not unduly affecting our measure. To further investigate, we now examine whether our primary test of construct validity is impacted by controlling for the authorizations process using our measures. To do this, we replicate Model 4 of Table 1 in the main text of the paper to examine its robustness to the inclusion of these variables. The results of this exercise are reported in Table 7. Overall, controlling for authorizations activity appears to have little impact on the inferences that we make with respect to the ally principle.

Models 1-3 include the natural logarithm of the number of laws passed in the agency's jurisdiction in the past year, three years, and five years, respectively. The estimated coefficients for these variables are substantively tiny and inconsistent in sign. They are also imprecisely estimated – none is statistically significant in any model. Taken together, this suggests little to no relationship between new legislative activity in an agency's jurisdiction and the level of discretion afforded to the agency in the appropriations process. This is consistent with the modest bivariate correlation results reported above. Furthermore, the estimated coefficient on *Chamber-President Disagreement* is highly stable across these three specifications. The two lagged versions of the Mayhew-based authorizations measure do reach statistical significance. We estimate that there is a positive cor-

Table 7: Ally Principle Models with Authorizations Measures

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Chamber-President Disagreement	-0.161**	-0.161**	-0.160**	-0.159**	-0.159**	-0.161**	-0.0970*
	(0.0401)	(0.0399)	(0.0402)	(0.0399)	(0.0396)	(0.0401)	(0.0474)
Ln(Laws Passed)	-0.00304						
	(0.0323)						
Ln(Laws Passed), 3 Yr.		0.0196					
		(0.0692)					
Ln(Authorizing), 5 Yr.			-0.00746				
			(0.0960)				
Ln(Mayhew Laws)				0.150			
				(0.126)			
Ln(Mayhew Laws), 3 Yr.					0.285*		
					(0.114)		
Ln(Mayhew Laws), 5 Yr.						0.337**	
						(0.103)	
Ln(Expired Programs)							0.118*
							(0.0574)
Inflation	0.0189	0.0179	0.0190	0.0187	0.0181	0.0173	-0.0283*
	(0.0126)	(0.0123)	(0.0122)	(0.0126)	(0.0122)	(0.0119)	(0.0113)
Unemployment	-0.00530	-0.00546	-0.00503	-0.00516	-0.00563	-0.00640	0.00135
	(0.0144)	(0.0143)	(0.0144)	(0.0145)	(0.0145)	(0.0145)	(0.0160)
Administration Change	-0.0519	-0.0496	-0.0503	-0.0505	-0.0504	-0.0514	-0.0579
	(0.0327)	(0.0302)	(0.0308)	(0.0306)	(0.0299)	(0.0295)	(0.0317)
War	-0.0714	-0.0708	-0.0707	-0.0707	-0.0683	-0.0659	0.0125
	(0.0483)	(0.0486)	(0.0498)	(0.0485)	(0.0478)	(0.0470)	(0.0391)
Public Approval	0.00452**	0.00451**	0.00449**	0.00449**	0.00456**	0.00431**	0.00404*
	(0.00159)	(0.00162)	(0.00168)	(0.00161)	(0.00162)	(0.00160)	(0.00179)
Agency FE	✓	✓	✓	✓	✓	✓	✓
Chamber-Subcommittee FE	✓	✓	✓	✓	✓	✓	✓
N	8284	8284	8284	8284	8284	8284	6174

Coefficients reported from OLS regression model, with robust standard errors clustered by agency in parentheses. Significance codes: *p < 0.05, **p < 0.01, two-tailed tests.

relation between these and total discretion, suggesting a long-run impact of significant legislation in the discretion levels for an agency. Notably, this suggests that if the effect of authorizations is on report lengths, then there is a negative correlation between activity and report length when controlling for the other factors in the regression (which would produce the observered positive correlation with discretion). With respect to the number of unauthorized programs in an agency's jurisdiction, we estimate a positive, and statistically significant relationship between discretion and unauthorized expenditures. Across all of these estimations, we find evidence for the ally principle that is consistent both in size and magnitude across all specifications.[7] Overall, we find that there does seem to be some modest relationship between the authorizations and appropriations processes, at least for the measure of appropriations discretion that we propose.

In sum, there does not appear to be any impact on our estimates of the key political variables in terms of magnitude or significance, however, when we account for this authorizations activity, suggesting that the activity itself is not a problematic omitted variable for our measure (at least in regard to the universe of analyses we explore in this paper). This implies that a standalone appropriations measure is, first, tapping into substantively distinct congressional decisions and, second, reflects the political dynamics of discretionary decisions that are the object of interest in theoretical and empirical studies of discretion. Overall, then, this analysis lends further credibility to using a standalone appropriations-based measure of discretion to test theories that implicate the concept.

---

[7]Note that the magnitude of the effect is slightly smaller in Model 7, with an estimated coefficient of -0.097. This is due not, we argue, to the inclusion of the expired authorizations but to the restricted years on with the analysis is conducted, i.e. only after FY 1987. If you restrict Model 4 in Table 1 to these years, the resulting estimated coefficient for divided government is also smaller than in the full sample of years, -0.12.

F Content Analysis of Committee Reports

In this section, we provide a content analysis of committee reports to further demonstrate the content validity of using their page length as a proxy for the constraints therein. To do this, we coded the report language for each agency in a 10% sample of our dataset for the presence of a number of constraints that have been commonly recognized in the literature on bureaucratic politics. In particular, we focus on five constraints previously identified as being among the most common and important ways that congressional actors can constrain the executive branch through statutory and appropriations language (Epstein and O'Halloran 1999; Potoski 1999; Potoski and Woods 2001):

- Reporting requirements that require agencies to provide information to Congress
- Deadlines that command agencies to agencies to take an action by a specific date or point in time (e.g. 60 days after legislation is passed)
- Additional requirements on agency reprogramming authority that limits their flexibility in using budget authority across functions
- Requirements that agencies study particular issues in order to inform the policy process
- Requirements that agencies consult with other actors prior to making policy or in the course of the implementation process

All of these requirements serve to limit the scope of agency action. While we argue that these are not the only types of constraints that we observe in committee report language (see Appendix A), if the presence of these constraints is correlated with the length of instructions given to agencies, then this should provide further evidence of content validity.

To code the reports, we relied on key word searches in order to identify the presence of each constraint. The search words we utilized for each type of constraint are listed in Table 8. Given that these words appear frequently throughout reports and are not always indicative of the presence

of a particular constraint, we relied on human coding to verify that each incidence of the key words was actually related to a constraint in order to prevent false positives.

In particular, we searched the PDF documents of each report for key words for each constraint category. For every hit produced, we read the corresponding page to determine whether it referred to a relevant constraint the committee is placing on an agency. We counted the number of constraints directed toward each agency in the report. To reduce the effect of outliers, we log these counts (after adding one) to produce a measure of each constraint by agency for each report in the analyses below.

Given the intensity in terms of time and labor of this task and the large number of agency-years in our dataset, we conducted this exercise on a random sample (10%) of all committee reports in our dataset between FY 1976 and 2015. This yields a sample of 80 distinct reports, and covers 951 agency-years. We divided the sample randomly between both coauthors to carry out this task.[8]

In order to assess the content validity of our measure of constraint, i.e. report length, we compare the correlations between the logged number of pages in the report for an agency and the logged number of each constraint that we coded in the reports. If these correlations are positive, then that suggests our measure of constraint is picking up on the incidence of these types of well-known and recognized constraints. We also created a constraint scale, summing all of the different constraints that we coded for each agency-report observation. The Cronbach's alpha for these measures of constraint is 0.94, suggesting a high degree of internal validity to such a scale. Additionally, we conducted a factor analysis and created a factor score of constraint that incorporates all five constraints that we coded.

To begin, we first examine the bivariate correlations between our measure of constraint (logged report length) and the incidence of the other constraints. Table 9, below, displays these correlations. As can be seen, all of the correlations are very large and positive, suggesting that our measure of

[8]An inter-coder reliability check on a sample of 33 agencies yielded a correlation of 0.99 between authors.

Table 8: Content Analysis Key Word Search Terms

Constraint	Search Term
Reporting requirement	Report
	Submit
	Plan
Deadline	Year
	Month
	Day
	January
	February
	March
	April
	May
	June
	July
	August
	September
	October
	November
	December
Reprogramming	Reprogram
	Notif
Required Study	Study
	Review
	Conduct
	Pilot
Consultation Requirement	Consult
	Partner
	Work with
	Collaborat
	Cooperat
	Coordinat
	Joint
	Work together

Table 9: Bivariate Correlation Between Report Length and Incidence of Constraints

Constraint Measure	Correlation with Ln(Pages)
Reporting Requirements	0.84
Deadlines	0.79
Reprogramming Requirements	0.64
Study Requirements	0.79
Consultation Requirements	0.74
Summary Measure	0.88
Factor Score	0.84

constraint is correlated with the presence of these other well-known constraints. All of the reported correlations are statistically significant with $p < 0.001$.

In addition to examining bivariate correlations, we also run a series of regression analyses so that we can control for agency and committee-year, which both may affect the report length and constraints. To do this, we regressed our measure of constraint onto the other measures we collected and included fixed effects for agency and committee-year. As can be seen in Table 10, even after controlling for agency and committee-year, we find positive and statistically significant relationships between the alternative measures of constraint and our page length measure of constraint.

Overall, these results provide strong evidence that the measure of constraint that we use throughout the paper is highly related to the presence of common constraints in the bureaucratic politics literature. This provides further support for the content validity of this portion of the measure.

Table 10: Content Validity: Report Pages and Individual Constraints

	Model 1	Model 2	Model 3	Model 4	Model 5	m6	m7
Reporting Requirements	0.796***						
	(0.0626)						
Deadlines		0.737***					
		(0.0693)					
Reprogramming			0.698***				
			(0.117)				
Study Requirements				0.850***			
				(0.0652)			
Consultation Requirements					0.808***		
					(0.0729)		
Summary Measure						0.712***	
						(0.0466)	
Factor Score							0.820***
							(0.0617)
N	950	950	950	950	950	950	950
Agency FE	✓	✓	✓	✓	✓	✓	✓
Subcommittee-Year FE	✓	✓	✓	✓	✓	✓	✓

The results reported in the table are from regressions of logged report pages for an agency on the logged number of particular constraints (or combinations, in the case of the summary measure and factor score). Estimated OLS coefficients reported with standard errors clustered by subcommittee-year in parentheses. Significance codes: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

G Comparisons to Previous Measures

In this appendix, we consider the statistical performance of our measure relative to others in the literature. In particular, we focus on the pioneering measure of Epstein and O'Halloran (1999), page-length based measures of discretion, and measures of discretion that are operationalized by the presence of a particular constraint (e.g. reporting requirements or deadlines). In addition to the conceptual critiques of these measures in the main manuscript, in this appendix we show that they are not always highly correlated with our measure, and all of them fail to reproduce the ally principle – our key metric of construct validity.

G.1 Comparison to EO Measure

First, we examine the relative performance of our measure and the EO measure of discretion. We were unable to obtain a replication file for Epstein and O'Halloran (1999), and here rely on data from MacDonald (2007), which created a similar measure of agency discretion using the CQ Almanac. To begin, we must determine whether the data from MacDonald (2007) sufficiently replicates the measure of discretion used in *Delegating Powers* in order to make comparisons between it and our measure.

MacDonald (2007) uses a modified version of the Epstein and O'Halloran measure of discretion based on his own coding of the CQ Almanac; however, it is possible to reconstruct their measure using his codings based on his replication files. We take two approaches to consider the similarity between the original EO measure and that which we constructed based on the MacDonald data. First, we consider the summary statistics reported by Epstein and O'Halloran to determine the degree to which these statistics match those from our reconstructed dataset based on MacDonald (2007). Then, we replicate Model 1 from Table 6.3 in *Delegating Powers*, which represents a key test of the ally principal in the book.

Table G.1 below compares the summary statistics (means and standard deviations) for the variables used in the Epstein and O'Halloran basic test of the ally principle. The left-hand column

reproduces the statistics reported in *Delegating Powers*. The right-hand column reports the statistics from our replicated measure of discretion and the covariates in the models. The standard deviations are presented in parentheses.

Variable	Delegating Powers Mean	Replication Dataset Mean
Discretion	0.18 (0.16)	0.19 (0.16)
Divided Government	0.55 (0.50)	0.55 (0.50)
Start Term	0.58 (0.50)	0.58 (0.50)
Activist	1.33 (1.51)	1.33 (1.51)
Budget Surplus	-0.075 (0.097)	-0.075 (0.097)

As can be seen, our replication dataset is nearly identical to that used in *Delegating Powers*, at least in terms of the summary statistics for the variables used in the ally principle analysis. The only discrepancy occurs for the measure of discretion, where our mean value is off by one hundredth. The standard deviation for the discretion measure is exactly the same as that of the original measure. This gives us confidence that our measure is at least extremely similar to that utilized in the Epstein and O'Halloran book, if not identical.

To further validate the measure, we replicated Model 1 in Table 6.3 of the book. Esptein and O'Halloran report that the estimated coefficient for the divided government variable from a robust regression was -0.032 with a t-value of -2.01. Our replication version of their measure in the same model with the same controls yields an estimated coefficient of -0.032 with a t-value of -1.82. In sum, we believe that the two measures are substantially similar. There are obviously some differences given the slight discrepancy in the mean of the discretion variable between the datasets and the estimated uncertainty in our regression relative to theirs. Nonetheless, since we are unable to use the original measure, we move forward with the replicated version in order to assess convergent validity with our measure, noting that it is not exactly the same as that used in *Delegating Powers*.

We now analyze the statistical performance of our measure in terms of construct validity relative to the EO measure. First, however, we note several important caveats. In general, it is difficult

for us to quantitatively assess the performance of our measure relative to other authorizations-based measures. The unit of analysis is wholly different for our measure relative to statute-based ones. For example, EO examine only a subset of significant legislation over a different time period than the one we examine. Further, there is no natural way to compare their measure to our appropriations-based measure because none of the legislation they examine corresponds to the appropriations legislation that we examine. Because the datasets are disjoint, traditional model comparison methods are inappropriate. Finally, our measure is based on a different data generating process than those that are authorizations-based, which clearly may influence comparisons.

To provide some statistical comparison, we reproduce the specifications of our ally principle test in Table 1 of the main paper using the EO measure. To do this, we needed to identify the agency receiving discretion on each piece of significant legislation included in their dataset, which was a challenging task given that most legislation dealt with several agencies at a time. As such, we identified a "primary" agency for each law based on the reading of the text and the associated policy area (see Appendix E for further detail). Figure 2 below, replicates Figure 4 in the main text using the EO measure and our coding of the "primary" agency. This, along with information provided by MacDonald concerning the responsible committee, allowed us to include agency and committee fixed effects analogous to those used in the analysis of our measure.

Table 11 replicates Models 1-4 of our ally principle analyses from Table 1 in the manuscript using the EO measure. Here, the unit of observation is the significant law (1947-1992), while ours is the chamber-agency-year. Hence, the number of observations is considerably lower in this analysis as compared to the models using our new measure. Overall, when accounting for agency differences both in terms of uncertainty estimates and the inclusion of fixed effects, divided government does not appear to be meaningfully associated with EO's measure of discretion. While the coefficient for divided government is in the expected negative direction, it is estimated with high levels of uncertainty. The R^2 for Model 4 in Table 11 is 0.57, while it is 0.82 for the comparable model using our measure in Table 1 in the main text of the paper. This suggests that these political variables that have been theorized and shown in some cases to impact discretion in the previous

Table 11: Epstein-O'Halloran Replication Results

	Model 1	Model 2	Model 3	Model 4
Divided Government	-0.0263	-0.0295	-0.0226	-0.0343
	(0.0243)	(0.0335)	(0.0321)	(0.0378)
Inflation				0.00212
				(0.00454)
Unemployment				0.00653
				(0.0156)
Administration Change				-0.0176
				(0.0434)
War				-0.0224
				(0.0366)
Public Approval				-0.000556
				(0.00142)
Major Agency Fixed Effects		✓	✓	✓
Committee Fixed Effects			✓	✓
N	256	256	256	256

Coefficients reported from OLS regression model using replication of the Epstein and O'Halloran measure of discretion, with robust standard errors clustered by significant agency in parentheses. Significance codes: *p < 0.05, **p < 0.01, two-tailed tests.

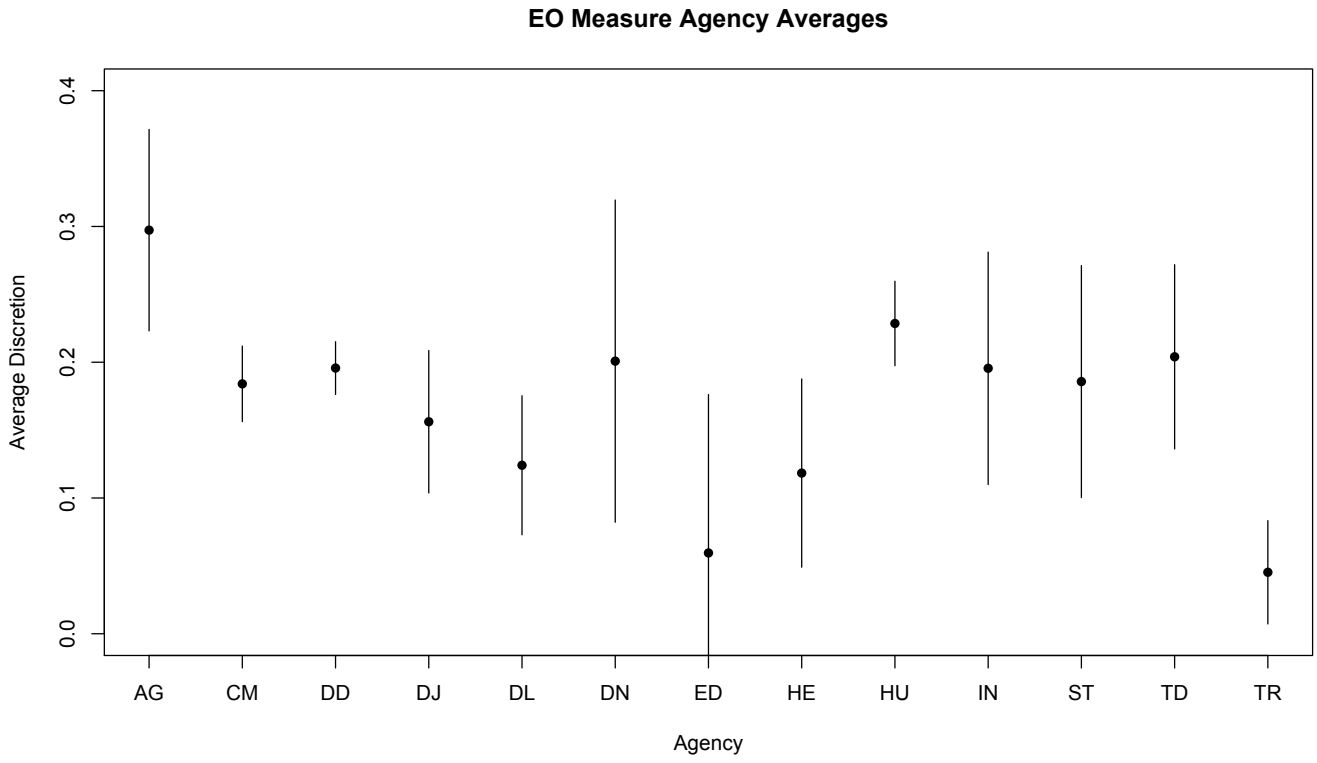

Figure 2: This figure displays the mean level of discretion for each Cabinet agency in the statutes where we determined the agency was the "primary" agency.

work also fit our data better than the EO measure.

One other potential complication in the comparison of these measures is that divided government used in Table 11 is different than that used in the main text of the paper. In particular, the variable in the EO analysis is system level divided government (i.e. at least one of the House or Senate is controlled by a different party than the presidency), whereas our unit of observation is at the chamber level. To ensure that this is not driving the results, we also examined our dataset using divided government at the system level to confirm that the results are unchanged. The estimated coefficient in the full model (i.e. analogous to Model 4) was -0.211, with a standard error of 0.042 ($p < 0.001$), ruling out the possibility that this difference is responsible for the different results.

Overall, this analysis using the EO measure reveals some insights. First, it demonstrates some practical limitations in examining discretion in authorizing statutes because it requires necessarily

subjective and potentially error-prone codings of major affected agencies. Relatedly, the measure cannot discriminate among different principals with the same ease as the appropriations-based one. Second, there is difficulty establishing the construct validity of the EO measure with the ally principle when taking into account agency differences. Finally, these analyses reiterate the utility of our proposed measure in overcoming these challenges. It is easy to identify agencies in the appropriations process and to associate the discretion granted to them to specific principals, while also replicating the ally principle across a range of specifications.

G.2 Page Length Measures

Another prominent measure of discretion in the literature is the page length of statutes (e.g. Clinton et al. 2012; Huber and Shipan 2002). The motivating idea behind these measures is that longer statutes are more constraining (and thus afford less discretion to the executive branch). However, as we discuss in the main manuscript, we believe that these measures do not capture delegation – an important component of discretion. Comparing these measures to our own is not completely straightforward, since we are not focused on authorizing statutes. However, in order to provide some comparison, we consider here how our measure compares to the length of appropriations bills.

To do this, we first aggregated our measure of discretion granted to agencies to the committee-year level. This allows us to create a bill-level measure of discretion. We then compare this aggregated version of our measure to the length of the bill that the committee reported, which is the natural analogue to our aggregated measure. The bills were located primarily using Congressional ProQuest, though a number were missing from the database, particularly from the 1980s. Altogether, we are able to compare discretion and bill length for 697 cases. In the analyses below, we use the natural logarithm of bill length given the presence of outliers. First, we assess the degree to which our aggregated measure correlates with bill length. Then, we examine the relative performance of our measure and the bill length measure in our core test of construct validity – the ally principle.

To begin, we find only a modest correlation between the aggregated version of our measure of discretion and appropriations bill length. In particular, the estimated correlation between the two measures is 0.3. Tellingly, we find a higher correlation between our measure of constraint (the logged number of pages in the committee report) and the logged bill length – 0.5. This accords with our idea that bill length is fundamentally tapping more into a notion of constraint, rather than discretion.

We also examined the performance of the bill length measures in the context of the ally principle. If bill length is operating as a measure of discretion, then we would expect to find a strong positive relationship between legislative-executive ideological divergence and bill length. In Table 12, we compare the ally principle results using our aggregated measure of discretion as well as the bill length measure of discretion. The analyses include fixed effects for the committee issuing the report/bill and also cluster standard errors at that level as well. We also include the same battery of control variables used in Model 4 of Table 1 in the main text of the paper.

Model 1 in Table 12 includes the results from a model regressing our aggregated measure of discretion onto the covariates described above. As can be seen, the familiar ally principle result obtains – preference divergence is negatively correlated with the amount of discretion afforded to the executive. The estimated coefficients in Model 2 are obtained by regressing logged bill length onto the same set of covariates. The estimated coefficient is very small, indicating that disagreement is associated with a 1.9% increase in the length of a bill. In addition to being substantively small, the measure is imprecisely estimated, with a standard error more than double the estimated coefficient.

In sum, the results here accord with the conceptual critiques of page length measures that we outline in the main text of the paper. In particular, we believe that bill length measures most readily map onto the idea of constraint. However, levels of constraint are not informative without some conception of the level of delegation in the bill as well. As both we and Epstein and O'Halloran (1999) demonstrate, delegation and constraint are often positively correlated. In line with these ideas, the bill length measure is only modestly positively correlated with our measure of discretion

Table 12: Measure Comparison: Page Length

	Model 1	Model 2
Chamber-President Disagreement	-3.682**	0.0191
	(1.586)	(0.0414)
Inflation	-0.252	-0.0720***
	(0.256)	(0.00709)
Unemployment	-0.801	-0.0489***
	(0.671)	(0.0135)
Administration Change	-0.846	-0.0101
	(1.210)	(0.0377)
War	-1.442	0.252***
	(1.517)	(0.0554)
Public Approval	0.0161	-0.00811***
	(0.0474)	(0.00197)
N	697	697
Measure	Aggregated Discretion	Bill Length
Chamber-Subcommittee FE	✓	✓

Coefficients reported from OLS regression models with robust standard errors clustered by committee in parentheses. Significance codes: *p < 0.10, **p < 0.05, ***p < 0.01, two-tailed tests.

but is more highly correlated with the constraint component of our discretion measure. Further, the measure fails to yield results that are consistent with the well-known ally principle. This is not just a function of smaller sample size either – our proposed measure yields a strong ally principle result even with this smaller sample. Overall, we believe that both the conceptual ambiguities of the page length measure as well as its poor statistical performance in a core test of construct validity cast doubt on the utility of page length as a measure of discretion.

G.3 Other Constraint Measures

In addition to page length measures, other potential measures of discretion focus on individual constraints. We believe that conclusions about discretion are not possible to draw from results that look only at constraints, for similar reasons articulated above. In particular, these measures do not in any way account for the degree of delegated authority to an executive actor. For instance, a law with one reporting requirement and five delegations of authority is not equivalent to a law with one reporting requirement and ten delegations of authority. In this section, we demonstrate support for this conceptual critique of constraint-based measures by comparing their correlations with our measure of discretion (as well as our measure of constraint) and also examining their performance with respect to the ally principle. In both cases, we believe these constraint-based measures fail to adequately capture the concept of discretion.

In order to explore these potential alternative measures, we use our content analysis of committee reports described in Appendix F, where we coded for the presence of five different constraints: reporting requirements, deadlines, reprogramming restrictions, study requirements, and consultation requirements. As was demonstrated there, these measures have strong correlations with our measure of constraint, i.e. report length. In Table 13, we examine whether those same correlations exist between our measure of discretion, our measure of constraint, and these five constraints. The correlation pattern is consistent across each of the five constraints. While the presence of each is highly correlated with the constraint component of our measure (i.e. committee page length), they are more modestly correlated with our measure of discretion. This again, reflects their limitations

Table 13: Correlations Between Constraints, Ln(Report Pages), and Discretion

Constraint Measure	Correlation with Ln(Pages)	Correlation with Discretion
Reporting Requirements	0.84	0.44
Deadlines	0.79	0.43
Reprogramming	0.64	0.31
Study Requirements	0.79	0.40
Consultation Requirements	0.74	0.38

as a measure of discretion, which includes both delegation and constraints.

In addition to these correlations, we also examine the individual constraint measures in terms of replicating the ally principle and compare the results to our measure of discretion. Table 14 displays the results of these analyses. We use the same specifications as are employed in Model 4 of Table 1 in the main text of the paper. Note that the number of observations is smaller in these samples because we only coded constraints for a 10% sample of the committee reports in our dataset, as detailed in Appendix F. The regression models below all use the same set of observations in each model.

As shown in Model 1, the results for our discretion measure mirror those reported in the main text of the paper. President-congressional ideological divergence has a large, negative, and statistically significant impact on discretion. The results suggest an 18% reduction in discretion during periods of partisan divergence. However, we do not see the same patterns for the individual constraint measures. In all cases, the estimated coefficients are substantively small and statistically insignificant. This suggests that these measures are not able to replicate the ally principle result, casting further doubt on their utility as measures of discretion.

Table 14: Measure Comparison: Constraints

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Chamber-President Disagreement	-0.196**	-0.0275	-0.0172	-0.00210	-0.0293	0.00557
	(0.0981)	(0.0427)	(0.0458)	(0.0292)	(0.0439)	(0.0533)
Inflation	-0.0269	-0.0116	-0.0198	-0.00615	-0.0141	-0.0264***
	(0.0253)	(0.0173)	(0.0162)	(0.0109)	(0.00936)	(0.00961)
Unemployment	-0.0697**	-0.00837	-0.0237	0.00434	0.000224	-0.0218
	(0.0347)	(0.0179)	(0.0150)	(0.0155)	(0.0127)	(0.0147)
Administration Change	0.146*	-0.165***	-0.105**	-0.0172	0.0110	-0.0123
	(0.0855)	(0.0523)	(0.0529)	(0.0370)	(0.0515)	(0.0452)
War	-0.269**	0.103	0.0900	0.0924	0.0112	0.0954
	(0.119)	(0.0888)	(0.0820)	(0.0725)	(0.0480)	(0.0610)
Public Approval	0.000696	-0.00132	-0.00182	-0.00103	0.0000651	-0.00226
	(0.00335)	(0.00239)	(0.00260)	(0.00155)	(0.00192)	(0.00204)
N	924	924	924	924	924	924
Agency FE	✓	✓	✓	✓	✓	✓
Chamber-Subcommittee FE	✓	✓	✓	✓	✓	✓
Measure	Discretion	Reporting	Deadlines	Reprogramming	Study	Consultation

Coefficients reported from OLS regression models with robust standard errors clustered by agency in parentheses. Significance codes: *p < 0.10, **p < 0.05, ***p < 0.01, two-tailed tests.

References

Bolton, Alex, Rachel Augustine Potter and Sharece Thrower. 2016. "Organizational Capacity, Regulatory Review, and the Limits of Political Control." *Journal of Law, Economics, and Organization* 32(2):242–271.

Carpenter, Daniel, Jacqueline Chattopadhyay, Susan Moffitt and Clayton Nall. 2012. "The Complications of Controlling Agency Time Discretion: FDA Review Deadlines and Postmarket Drug Safety." *American Journal of Political Science* 56(1):98–114.

Clinton, Joshua D, Anthony Bertelli, Christian R Grose, David E Lewis and David C Nixon. 2012. "Separated Powers in the United States: the Ideology of Agencies, Presidents, and Congress." *American Journal of Political Science* 56(2):341–354.

Davis, Otto A., M. A. H. Dempster and Aaron Wildavsky. 1966. "A Theory of the Budget Process." *American Political Science Review* 60(3):529–547.

Epstein, David and Sharyn O'Halloran. 1999. *Delegating Powers: A Transaction Cost Politics Approach to Policy Making under Separate Powers*. Cambridge University Press.

Fenno, Richard F. 1966. *The Power of the Purse: Appropriations Politics in Congress*. New York, NY: Little, Brown.

Fisher, Louis. 2005. Legislative Vetoes After Chadha. Technical Report Order Code RS22132 Congressional Research Service http://www.loufisher.org/docs/lv/4116.pdf: .

Howell, William G and Saul P Jackman. 2011. "Inter-Branch Bargaining over Multiple Outcomes." *American Journal of Political Science* 57(4):956–970.

Huber, John D and Charles R Shipan. 2002. *Deliberate Discretion?: The Institutional Foundations of Bureaucratic Autonomy*. Cambridge University Press.

Jones, Bryan D., László Zalányi and Péter Érdi. 2014. "An Integrated Theory of Budgetary Politics and Some Empirical Tests: The U.S. National Budget, 1791-2010." *American Journal of Political Science* 58(3):561–578.

Kiewiet, Rodrik and Matthew McCubbins. 1991. *The Logic of Delegation*. Chicago University Press.

Lewis, David E and Jennifer L Selin. 2012. *Sourcebook of United States Executive Agencies*. Washington, DC: Administrative Conference of the United States.

Light, Paul Charles. 1999. *The True Size of Government*. Washington, DC: Brookings Institution Press.

MacDonald, Jason A. 2007. "Agency Design and Postlegislative Influence over the Bureaucracy." *Political Research Quarterly* 60(4):683–695.

Mayhew, David R. 2005. *Divided We Govern: Party Control, Lawmaking, and Investigations, 1946-2002*. 2 ed. Yale University Press.

Padgett, John F. 1980. "Bounded Rationality in Budgetary Research." *American Political Science Review* 74(2):354–372.

Potoski, Matthew. 1999. "Managing Uncertainty through Bureaucratic Design: Administrative Procedures and State Air Pollution Control Agencies." *Journal of Public Administration Research and Theory* 9(4):623–640.

Potoski, Matthew and Neal D Woods. 2001. "Designing State Clean Air Agencies: Administrative Procedures and Bureaucratic Autonomy." *Journal of Public Administration Research and Theory* 11(2):203–222.

Schick, Allen. 2008. *The Federal Budget: Politics, Policy, Process*. Brookings Institution Press.