Measuring Change and Influence in Budget Texts

Devin Judge-Lord
JudgeLord@Wisc.edu

August 30, 2017

Early-stage Outline for Discussion at APSA 2017 Only

This paper introduces a dataset of over 100,000 pages of discussion about how U.S. federal budget line items ought to be used and explores how well a variety of methods commonly used to identify similarity, difference, and influence among texts travel to this new context. Ultimately, it aims to test theories about the role of the executive branch in influencing the policy agenda for congressional appropriations and the role of Congress in influencing the policy agenda of executive agencies. This involves two methodological exercises. First I contrast several approaches to measure similarity and differences in policy texts were we have strong intuitions about their relationships. For example, we expect presidential transitions to affect the content of the president's annual budget. Different text analysis methods give us different perspectives on the magnitude and substance of change. Second, it outlines an approach for estimating the relative influence of texts on a change in a policy document, for example, the influence of the president's budget on congressional appropriations committee reports and vice versa.

1 Introduction

On July 1, 2016, the Director of the Obama Administration's Office of Management and Budget circulated a short memo to all federal agencies: "The FY 2018 Budget will be submitted by the next President, and agencies are not required to submit a formal budget request in September" (OMB 2016).

The allocation of funds is widely recognized as an important part of politics and has been well studied in political science. Equally important, but much more difficult to study is what government agencies do with these resources. The short answer is that government does too many things to properly catalog and many specifics are unknown at the time the budget is proposed. Nevertheless, one place to start is to look at what the executive branch says it will do with the funds requested and what appropriations committees in Congress say the funds are for. Who drives this more qualitative aspect of budgeting and what can it tell us about executive and legislative power?

An agency budget justifications can be read as a kind of manifesto, a comprehensive description of priorities, value propositions, and specific goals. Each agency, guided and overseen by the White House

Office of Management and Budget, submits several hundred to several thousand pages alongside their request to Congress each year. Congressional appropriations subcommittees that oversee appropriations to that agency respond to the president's budget request with their own report outlining priorities and instructions to agencies.

From the executive side, what agencies highlight in their budget requests can be seen as a statement of presidential or agency priorities, which may or may not align. For example, the fiscal year 2018 Budget Justification for the the Environmental Protection Agency (EPA) reports achievements in addressing climate change while simultaneously announcing the elimination of many climate related programs, overall using the term "climate change" 91% less than the previous year.

Figure 1: "Climate Change" in the Environmental Protection Agency Budget Justification

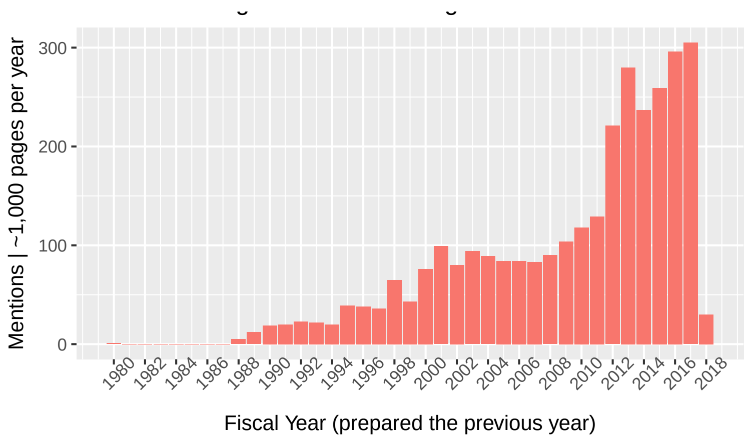
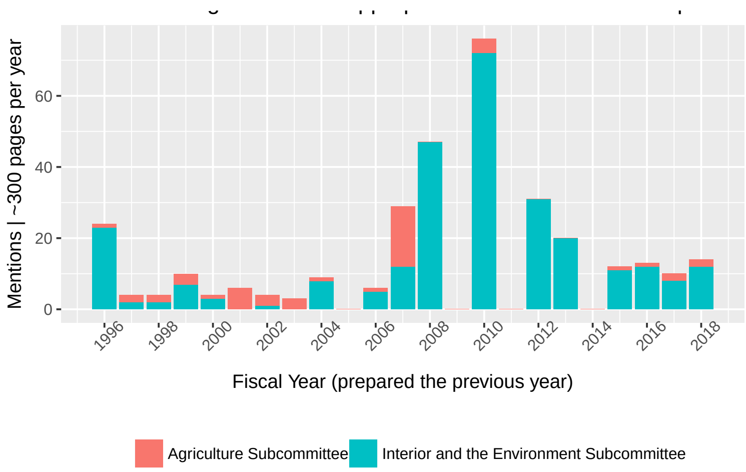


Figure 2: "Climate Change" in House Appropriations Subcommittee Reports



On the legislative side, the detailed reports accompanying appropriations bills reflect the priorities of committee members and the majority party. For example, discussion of "climate change" increases after the Democrats won a majority in 2006 and decreases after the Republicans won a majority party in 2010.[1]

Yet variation any one term is far from an overall picture of stability or change. From figures 1 and 2 it is also unclear if shifts in presidential priorities influence Congress or if shifts in congressional priorities influenced the EPA's budget requests. Assessing change in issue attention and agenda setting will be the aim of this paper.

We may expect that what an agency proposes to do with its budget affects appropriation of funds by Congress. Conversely, agencies may expect that the detailed appropriations committee reports that accompany their budgets are what Congress expects them to do. Figure 3 illustrates these intuitions as Hypotheses 1 and 2.

For the purpose of developing measures of influence, the iterative structure of annual budgeting between the president and Congress offers an especially straight forward context. Generally speaking, each agency budget justification aims to influence policy decisions reflected in one pair of appropriations subcommittee reports and each appropriations report aims to influence policy decisions reflected in one budget justification. The president's fiscal year 2018 budget proposal aims to influence the outcome of the 2018 appropriations bills (described in the fiscal year 2018 appropriations subcommittee reports) and each subcommittee report aims to influence agency actions and priorities that will be described in its fiscal year 2019 budget justification.

These data also present an intuitive way to test potential measures of document similarity and change. Because policy priorities reflect party control, measures ought to capture dis-continuities caused by changes in party control.

This paper utilizes 10 or more years of budget justifications from 70 federal agencies falling under the authority of three departments (the US Department of Agriculture, Department of Health and Human Services, and Department of the Interior) and one independent agency (the Environmental Protection Agency). These agencies constitute the bulk of the jurisdiction of two Senate and two House appropriations subcommittees (the Agriculture, Rural Development, Food and Drug Administration, and Related Agencies Subcommittees and the Interior, Environment, and Related Agencies Subcommittees). Thus, for each year 2008-2017 there are roughly 71 agency documents and exactly 4 appropriations committee reports. Documents range from several dozen to several thousand pages, most being between 100 and 600 pages. The result is well over 100,000 pages of discussion about how federal funds ought to be used.

1. The House Appropriations Subcommittee on Interior, Environment, and Related Agencies did not publish reports in 2005, 2009, and 2011.

These data may shed new light on at least four things of interest to political scientists: (1) the extent to which agencies or the president set the congressional agenda in budgeting, (2) how responsive agencies are to Congress, (3) how much presidential transitions, party control, and committee membership matter for policy content, and (4) whether congressional attention (or lack thereof) to an issue or agency is a good or bad sign for the size of corresponding appropriations. This paper focuses on the first three with respect to the textual recored.

2 Attention Allocation and Agenda Setting

Citing a *New York Times* article titled "U.S. Research Lab Lets Livestock Suffer in Quest for Profit," the House Appropriations Committee[2] recommended cutting Agriculture Research Service (ARS) appropriations by $10 million even though ARS had requested a $60 million increase. Additionally, the Committee recommended withholding 5 percent of ARS's budget until it had addressed the Committee's concerns. This is bad attention. In contrast, the Committee had mild praise and a few polite suggestions for the Natural Resources Conservation Service (NRCS), mostly the same as the previous year. NRCS asked for a $10 million increase and received $5 million. Committee members may be paying some attention to NRCS, but not too much. Finally, the Agricultural and Plant Health Inspection Services (APHIS) received nearly twice as much attention (i.e. twice the number of new sentences) as it had the previous year and a budget increase twice as large as well. This included detailed instructions for what to do with the additional funds and praise for the importance of key agricultural constituents like citrus growers and the agency's efforts to serve them, even when "not completely successful."

These short examples highlight several nuances of issue attention. It is not surprising that representatives from agricultural states would attend to emerging issues of plant and animal diseases raised by the ARS and APHIS. Yet, the issue that got the most attention was not the failure of APHIS to stop the parasitic broomrape, but animal welfare in ARS research facilities, an issue raised more by PETA and the *Times* than the agency. In this instance, it was the *Times* and not the agency that got the Committee's attention, leading the committee to exercise "fire-alarm" control, a bad kind of attention from the agency's perspective. These examples show that attention varies from year to year, even where party control is constant and funding is relatively non-partisan.

In contrast to the praise heaped on APHIS and alarm for animal welfare toward ARS, the committee

2. These examples of different types of attention from the 2016 and 2017 House Appropriations Subcommittee on Agriculture, Rural Development, Food and Drug Administration, and Related Agencies Committee Report.

report section on NRCS only changed a few sentences from the previous year. This is striking in context of what NRCS was asking for. In the purpose statement on the first page of its 2017 budget justification NRCS writes that its "primary focus is to ensure that private lands are conserved, restored, and made more resilient to environmental challenges, like climate change." At the time the 2017 fiscal year report was published, there was little support in Congress for addressing what the House Subcommittee Chair called "Climategate." The budget report eschews the *climate* issue frame, which was raised 69 times in the NRCS budget justification. But instead of chastising NRCS as the Committee had chastised the Environmental Protection Agency for prioritizing climate change, the Committee commends the efforts of NRCS to mitigate increased flooding and recognizes the importance expanding irrigation to places that have not previously needed it. This highlights the advantage of focusing on issues (collections of words including flood and drought) rather than single words such as climate and how the impact of disagreement over issue frames on a budget may be conditioned on the underlying agreement on the issue.

Much attention has been paid to the agenda setting power of the president's budget (Wildavsky 1964; Brady et al. 2016; Whittington and Carpenter 2003; Fisher 2015; Berry et al. 2010). Like the above quote from the former Office of Management and Budget Director, this literature leads us to expect that budgets will highlight different priorities under new administrations. Yet the budget blueprints that originate in the White House are sparse. Agencies prepare their budget justifications and the Office of Management and Budget reviews and amends them to the extent the White House is attentive and able to amend. Budget justifications should be seen as a mixture of the priorities of presidential administration and career agency managers.

I hypothesize that what the executive branch asks for in discretionary budgeting influences Congressional attention in appropriations and that, in turn, agencies respond to congressional priorities. We know that agencies and their allies lobby Congress effectively (Carpenter 2001; Carpenter 2014), but budgeting and committee budget reports are also seen as mechanisms of congressional control (Yackee and Yackee 2009; Bolton and Thrower 2015; McCubbins and Schwartz 1984). We also know that the attention of policymakers over any set of issues is limited (Jones and Baumgartner 2005). A cursory reading of appropriations committee reports aligns with this scholarship. They appear to be partially paraphrasing agencies' own justifications for their budget requests, partially constraining budgetary discretion, and partially copied from the previous year's report. I thus expect these congressional budget reports to discuss a topic more or less when agencies do so first. I also expect agencies to respond to what Congress instructed them to do the previous year. Figure 3 shows these two hypotheses: ($H1$) appropriations committee reports tend to shift emphasis from year to year in the same direction as agency budget justifications and ($H2$) agency budget justifications tend to shift their emphasis in the direction that the appropriations committee did the previous year.[3]

Figure 3: Agenda Setting

$$
\begin{array}{cccccc}
\longrightarrow & \boxed{\text{Agency Request}_t} & \longrightarrow & \boxed{\text{Agency Request}_{t+1}} & \longrightarrow & \ldots \\
 & \downarrow (H1) & \uparrow (H2) & \downarrow & \uparrow & \\
\ldots & \longrightarrow & \boxed{\text{Committee Report}_t} & \longrightarrow & \boxed{\text{Committee Report}_{t+1}} & \longrightarrow
\end{array}
$$

Theories that focus on positive agenda control find that coalitions compete for the attention of policymakers, suggesting that attention is often good for their issue (Baumgartner and Jones 1991; Jones and Baumgartner 2005; Kingdon 1995).[4] Conversely, principal agent theories of Congress and the bureaucracy (e.g. "fire-alarm-control") suggest that attention often means sanction, opposition, or constraint (McCubbins and Schwartz 1984; Bolton and Thrower 2015).

Scholarship on Congressional committees and appropriations has focused on party agendas, appropriations to districts, and the politics of attention.

While the act of passing a budget may be seen as a fairly non-ideological sign of party competence (Butler and Powell 2014; Lee 2016), the content of a budget report likely reflects party and committee member agendas (Shepsle and Weingast 1987; Adler and Lapinski 1997; Lee 2000). Even if achieving the party agenda is seen as a general sign of competence (Cox and McCubbins 2005), the content of that agenda is ideological and we expect committees to pay more attention to issues they care about. Yet Berry et al. (2010) find that only some program budgets are affected by partisan control in Congress. Importantly, like Lee (2000), they do not find that committee members drive program spending toward their own districts. My data allow new tests of the extent to which committee chairs and partisan control matter by looking at the text added, deleted, and edited on issues where the old and new committee members disagree. We may see partisan control effects, committee chair effects, both, or neither. I expect committees to attend to issues raised in agency budget justifications proportionally to the chair's support for each issue.

Corollary to the politics of attention is the politics of inattention. Members of Congress lack the time to read many bills (Curry 2015), much less thousands of pages of budget justifications from administrative

3. Agency budget justification texts are indexed to the year t that the corresponding appropriations committee report is published, i.e. the year the budget is passed and the year before the fiscal year it funds. Agency budget justifications are published between 9 and 18 months before the fiscal year begins. Appropriations reports are published prior to the budget going to the floor, generally 2 to 6 months before the fiscal year begins.

4. Groups may also organize to advocate for a part of the budget to be cut or to restrict what funds can be used for, but most scholarship points to examples of groups organizing to push policymakers to attend to problems that require more spending on their issue rather than less. Jones and Baumgartner (2005) find that attention is more punctuated around budget and policy expansion than retrenchment. In the specific case of budget justifications, it seems unlikely that agencies frequently advocate for smaller budgets.

agencies. It may not be surprising that a substantial amount of text that appropriations committees send to the floor is copied either from the previous budget report or agency justifications. The stable core of discretionary budgeting likely represents non-salient issues that are simply ignored due to the limits of information processing (Jones and Baumgartner 2005) or issues of broad agreement (Lowi 1967; Adler and Wilkerson 2012). Additionally, industry groups target relevant committee members of both parties with campaign contributions in order to get their attention to industry-specific problems and redirect government spending in their direction (Powell and Grimmer 2016). The stable core of congressional budget justifications that I observe could thus represent low-salience issues, bipartisan issues, and issues on which committee members of both parties are consistently captured by industry.

In contrast, scholarship on the bureaucracy and Congress has focused on budgeting, and appropriations committee budget reports specifically, as a mechanism of sanction and constraint on delegation.

Wildavsky (1964) proposed a model of the budgeting process relating to how much agencies ask for, how much Congress allocates, and the strategies actors employ to get what they want. More rigorous modeling in the law and economics tradition takes seriously that bureaucrats may have political agendas as well as resource objectives. McCubbins et al. (1987) offer a framework with two general types of control: oversight and administrative procedures. Budgeting is seen a form of oversight. Yackee and Yackee (2009) use the discretionary share of budget as a measure of the strength of an agency's relationships with elected officials. Bolton and Thrower (2015) use the length of committee budget reports as a measure of how much Congress constrains different agencies.[5] Bendor et al. (2001) suggest that rational principals will delegate to agents with similar goals, repeated interactions, and when they are able to overcome commitment and information problems.

The repeated interactions of annual budget cycles has been a core subject of principal agent scholarship on bureaucracy. This scholarship assumes that budget authority is a form of discretion and that accompanying texts like appropriations committee budget reports are a form of constraint, but this relationship remains largely untested. Importantly, because of the focus on text as means of constraint,

5. Bolton and Thrower (2015) use the length of these committee reports relative to the size of the budget to measure how much Congress constrains agencies. Using this as a measure of executive discretion, they find that Congress gives greater discretion to ideologically aligned presidents. However, my finding that committee budget reports are highly stable from year to year (even more stable in length than in content), suggests that this is primarily driven by more funding going to issues on which Congress and the president agree while the length of budget reports stays the same. By looking at how the words of these reports change rather than just their length, we may be able to better measure the relationship between budget appropriations and these texts. Furthermore, as Fisher (2015) notes, committees have been know to sanction agencies for using funds for purposes not specified in their budget justification. This aligns with my intuition that sometimes additional attention from Congress is beneficial (e.g. when an agency adds to the scope of its request).

By measuring how agency and committee justifications align, we can know when a committee is expanding and when it is constraining what an agency can do with its budget, giving scholars a broadly useful new measure of the discretion granted to executive agencies by Congress. For example, studies of bureaucratic policymaking could benefit from this policy-specific measure of congressional opposition.

the assumption in this literature is that attention from Congress in these texts bodes poorly for agency budgets.

The next section describes several ways to approach the problem of assessing similarity and change across texts and reports intermediate results. Section 4 discusses potential next steps.

3 Multiple Measures of Similarity, Change, and Influence

Measuring similarity or change in policy and measuring influence across actor types are common objectives in political science. The structured nature of budget texts offers a testing ground for measures of similarity, difference, and influence among policy texts. In this case, both the independent and dependent variables are texts and the objective is to identify a set of feature or statistics that provides the most leverage.

Text reuse methods and term frequency comparisons or clustering methods have been used to measure nominally similar but conceptually distinct phenomena. For example, identifying the origin of an idea (as measured by either a string or a distribution of words) relies on textual evidence. As Wilkerson et al. (2015) note, a "policy idea" is "an admittedly ambiguous concept. For some, policy idea refers to a general policy objective (e.g., universal health care), whereas for others it refers to specific policy provisions in laws" (p. 945). The appropriate strategy depends on the nature of the texts and the nature of the idea under study, for example how much information depends on the ordering of words or their likelihood of appearing. In policymaking (i.e. when a policy text is the dependent variable), the conceptual difference between clustering method and text reuse methods may often map onto the distinction between political issues or issue frames and specific policy provisions or proposals. In broader political contexts, text reuse can also represent the repetition of political rhetoric, which is more helpful in identifying the influence of issue salience or agreement than specific causal relations like the origin of ideas.

Both topic modeling and text reuse methods have been used to study the "origins" of politically-relevant ideas. Brookhart and Tahk (2015) trace the origins of ideas using a dynamic topic model to discover whether topics were first raised by elites, the media, or the public. For them, an idea is a policy issue, usually a policy problem, represented by the frequency with which certain words appear. The same idea can be expressed a number of ways, but generally involves the a similar distribution of words that is distinct from the words one may use to speak about a different policy issue. They give the example of “human trafficking” which could also be described as “trafficking of sex workers,” both of which capture a similar underlying policy problem. This is appropriate because they aim to discover latent issues or issue frames in the frequency of word use. Focusing on matching sequences of words would fail to associate many texts that are about the same issue without copying text verbatim. Copied text may be interesting in this context. Indeed, when politicians or Twitter users quote new stories, this is evidence of the influence of the media, but quoted text would only identify the origins of the specific a phrase or fact, not the origin of the issue or issue frame.

In contrast, Wilkerson et al. (2015) uses text reuse to trace the origin of policy ideas in legislation. In this context an “idea” is a specific policy solution, an exact string of text crafted to address some aspect of a policy problem. Unlike the news media and Twitter content used by Brookhart and Tahk, there are not norms against copying text from others. Indeed, it is common. Rewriting legal language from scratch is more costly than rephrasing a tweet and policy ideas are more easily traced as an exact string of words rather than a distribution. Wilkerson et al.’s unit of analysis is a specific statutory provision. Because it is plausible that statutory provisions may have very similar distributions of words, but very different legal meanings and because statutory language is, in fact, often copied, using text reuse allows for a more nuanced measure. There is some risk of failing to identify earlier variations of the policy idea that used different wording, but given the norm of copying existing text, if a text is not copied, it may be reasonable to infer that it represents something new.

Each of the approaches reviewed below captures a different kind of similarity or difference between text.

Many existing methods use word frequencies to characterize the document similarity. Word frequencies may reflect policy priorities. For example, the frequency of the word “climate” reflects attention to the issue of climate change, which decreased sharply in the 2018 budget. I focus on three methods based on word frequency: cosine similarity, dimensional scaling with *Wordfish*, and topic models.

Text reuse may provide information about the attention that members of Congress devote to issues that overall topic proportions do not. For example, text that is copied from the previous year may indicate issues that received less attention than those where new text has been added or edited.

Hertel-Fernandez and Kashin (n.d.) use text reuse to detect the proposal and adoption of legislative language proposed by the American Legislative Exchange Council (ALEC), a group that drafts model state legislation and advocates for it. They find that states with less legislative professionalism are more likely to introduce and pass bills that contain ALEC language. Because they have external information about the document generating process—they know that ALEC plays a major role in writing and disseminating model legislation to state legislators—Hertel-Fernandez and Kashin (n.d.) can infer that the similarities they find are evidence of ALEC influence. Even if some state legislators are copying other states, ALEC is influential if it is the originator *or* disseminator of the policy text, a plausible assumption for many of these texts.

Recent research has begun to incorporate information about the network of relationships among actors, greatly improving the plausibility that text similarity reflects influence. Linder et al. (2017) find that state legislators with similar voting behavior introduce bills with matching text and that text reuse reflects policy diffusion networks found by other scholars. Garrett and Jansa (2015) assess the relative influence of interest groups and early adopting states on policy diffusion using text similarity as an attribute of a network model.

Question	Approach	Metric
How much discussion of K topics?	Topic Models	Proportions of K topics per document
	↳ Mixed Member	Topic proportions
	↳ Single Member	Number of paragraphs per topic
How similar is overall word-use?	Cosine Similarity	Similarity score for each pair
What are documents' relative positions?	Scaling (e.g. *Wordfish*)	Location on 1-Dimensional scale
What is copied from where?	Text Reuse	Tokens in document d copied from d'
	↳ Sentences	Matching sentences
	↳ 10-grams	Matching sequences $\geqslant$ 10 words
	↳ Smith-Waterman	Approximately aligned sequences

None of the above methods perfectly captures either the type of variation we aim to observe between versions of policy documents or between policy documents and potential sources of influence. Each captures a different kind of similarity or difference. As the next section shows, none seem to perform well in practice for the case of budgeting.

3.1 Selecting and Preprocessing Texts

As each committee oversees more than one agency, the first step was to select the relevant portions of committee reports. I select all pages that contain the agency's name or abbreviation.

The potential impact of preprocessing decisions with respect to each corpus (i.e. each collection of an agency's budget justifications and corresponding pages of appropriations reports) was assessed using the PreText() package (Denny and Spirling 2017). PreText results (reported in the appendix) suggest that removing stopwords and punctuation and using ngrams may be desirable. Indeed, in the context of budget documents the differences in documents due to punctuation appears to be systematic differences in formatting between agencies and committees that should be ignored. Thus, I remove punctuation when tokenizing text. Due to computational limits, I do not use n-grams for topic models in this draft. Removing numbers and converting to lower case have low estimated risk for topic model results and are helpful for identifying matching strings in texts often punctuated by budget numbers that change from year to year. These steps so are done to all texts. Stemming and covering text to lowercase are only done for term frequency based methods.

3.2 String Matching: Sentences, Smith-Waterman Alignment, and N-grams

Sentences are natural lexical units that may often be copied in their entirety. In Figure 9, "Sentences" is simply the percent of sentences in a document that have an exact[6] match in the corresponding document from the previous year.

Another way to identify copied, cut, and added text is the Smith-Waterman algorithm (Smith and Waterman 1981; Wilkerson et al. 2015). Because multiple aligned sections of documents may not appear in the same order, alignment must be identified at the paragraph or sentence level. Figure 4 shows one such local alignment between the fiscal year 2017 and 2018 reports of the House Appropriations Subcommittee on Agriculture, Rural Development, Food and Drug Administration, and Related Agencies.

Figure 4: Sentence-level Smith-Waterman Local Alignment

FY 2017:

```
 "When the Subcommittee heard from the FDA it focused on preventing burdensome
regulations for producers and the American people in addition to ongoing discussions
of how the FDA is implementing the Food Safety Modernization Act FSMA the XXXXXXXXX
motivation XXXXXX for XXXXX the generic drug labeling rule and regulation of tobacco
products"
```

FY 2018:

```
 "When the Subcommittee heard from the FDA it focused on preventing burdensome
regulations for producers and the American people in addition to ongoing discussions
of how the FDA is implementing the Food Safety Modernization Act FSMA XXX addressing
XXXXXXXXX opioid XXX abuse XXX XXXXXXX XXXX XXXXXXXX XXXX and regulation of tobacco
products"
```

The Smith-Waterman algorithm identifies an optimal local alignment of all words, identifying specific words and strings that were added (e.g. "the, motivation, for, the generic drug labeling rule") or cut (e.g. "addressing, opioid, abuse").

Similar to whole-sentence matching described above, Smith-Waterman can be used to identify the

6. Because punctuation and stopwords are removed, "exact" matching of sentences overlooks some minor edits.

Figure 5: Moving 10-Gram Matching Window

FY 2017 → FY 2018:

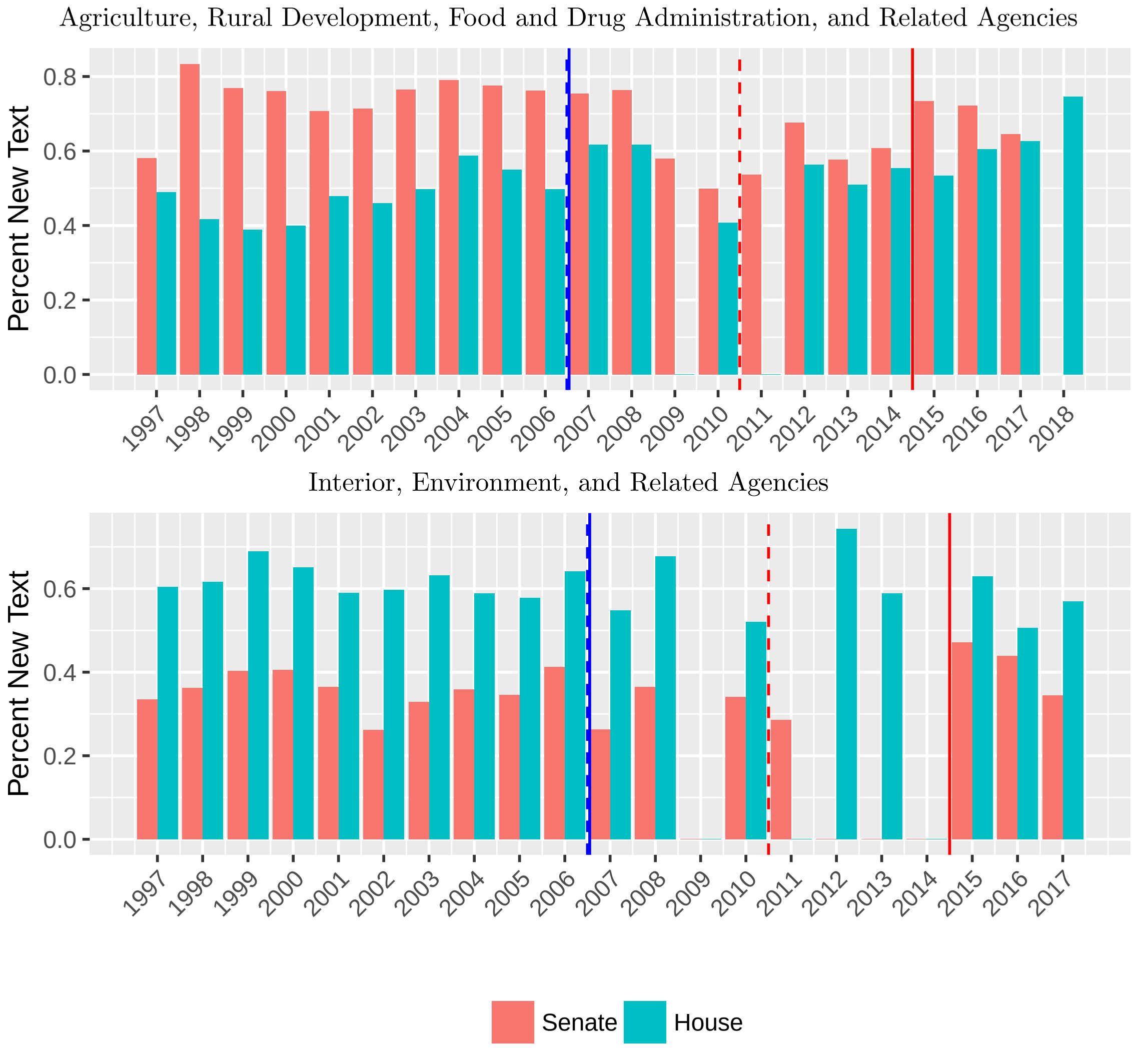
''This funding change eliminates climate change research FTE This funding change reduces air quality research This eliminates funding for the Science to Achieve Results STAR program for FY Statutory Authority:Clean Air Act Title II of Energy Independence and Security Act of Environmental Research Development and Demonstration Authorization Act ERDDAA Intergovernmental Cooperation Act National Environmental Policy Act NEPA Pollution Prevention Act PPA Global Change Research Act"

best match by selecting sentence or paragraph with the highest alignment score. However, because we are no longer selecting exact matches, nearly every sentence will contain strings that can be found in other documents. Selecting a minimum string length can reduce matches occurring by chance, for example, by only counting matches with a matching string of 10 words.

Casas et al. (2017) suggest a more straightforward approach of a "moving window" for identifying matches. They find a 10-word string to be an optimal size window for identifying instances of copying in policy text. Matches can be identified by tokenizing documents into all possible 10-gram strings, identifying matching tokens across documents. The vector of matches index all of the words that begin a matching 10-word sequence. This approach fails to capture short matching sequences where words have been added, edited, or deleted, but it has the advantage of capturing alignments across multiple sentences or paragraphs.

Figure 5 shows a section of the House Appropriations Subcommittee on the Interior, Environment, and Related Agencies where a match (underlined) would not have been identified by sentence-based matching but is captured by the 10-gram moving window. In this case, the copied text is simply a list of statutory authorities which are being used to justify the exact opposite policy approach from the previous year. Such boilerplate text may lead methods based on word frequency to find less difference among texts than we might otherwise expect. Figure 6 shows the percent of "new" (i.e. not copied) text appearing in each appropriations reports using this method.

Figure 6: Change in Appropriations Reports with a 10-gram Window



*Solid lines indicate change in Senate majority, dotted lines a change in House majority. Years where no report was published appear as 0. Differences are calculated with respect to the last year a report was published.

3.3 Methods Using Term Frequency

3.3.1 Cosine Similarity and Dimensional Scaling

Cosine similarity measures the frequency with which terms are used between pairs of documents. These scores are calculated for each corpus and normalized between 0 and 1. Two documents that use the same words with the same frequencies would have a cosine similarity score of one. Two documents that use none of the same words would have a score of 0. To convert this to a measure of year-to-year difference (i.e. the kind difference we expect to see with a change in party power), I use 1 - *normalized cosine similarity*. In figure 6 Cosine "difference" is the score between the a document and the corresponding document from the previous year.

Another way to use word frequencies to describe the relationships between texts is dimensional scaling. Building on work by Laver et al. (2003), Slapin and Proksch (2008) and others have used the *Wordfish* method of scaling to a single dimension to estimate the relative positions, ω, of political texts over time. To locate documents in policy space, word frequencies are assumed to be generated by a Poisson distribution. *Wordfish* scores shown in figure 5 are the absolute value of the differences in *Wordfish* scores between a document and the corresponding document from the previous year, $|\omega_{t+1} - \omega_t|$, normalized between 0 and 1.[7]

3.3.2 Topic Proportion

As noted, two classes of topic models may be used to estimate the proportion of documents devoted to a given topic. If documents are parsed by paragraph or sentence, a single-member topic model can classify each. Topic proportions would then be the percent of words belonging to blocks assigned to each topic. Alternatively, a mixed membership model can estimate topic proportions. Results reported here are based on the *Latent Dirichlet Allocation* (LDA) model described by Blei et al. (2003) estimated with variational expectation maximization and initialized with several passes of collapsed Gibbs sampling as described in Roberts et al. (2017)[8].

7. Cosine similarity and *Wordfish* scores were calculated using the quanteda package (Benoit et al. 2017). Matching sentences were calculated by tokenizing each document by sentence using the quanteda::tokenize() function. Ten-gram matches used a string matching function to identify if the same ten-gram sequence appeared anywhere in each text.

8. Results in this draft are estimated in LDA with common α and β priors and uncorrelated topic distributions. There were no obvious differences in results estimated with the Structural Topic Model (STM) where topic distributions are a function of the interaction of document author and year. Instead of covariates only being used post-hoc (estimating effects *after* naively estimating topics), STM brings information contained in covariates into the topic model by (1) assigning unique priors by covariate value, (2) allowing topics to be correlated, (3) allowing word use within a topic to vary by covariate values. Instead of $Pr(\theta) \sim Dirichlet(\alpha)$, topic proportions can be influenced by covariates X through a regression model, $Pr(\theta) \sim LogisticNormal(X\gamma, \Sigma)$. This helps the model avoid having to develop a categorization scheme from

Several potential approaches to measure political influence emerged from Blei et al. (2003) *Latent Dirichlet Allocation* (LDA) mixed-membership model. For example, several studies using structural topic models have aimed at measuring influence by measuring treatment effects on survey experiment responses (Roberts et al. 2014; Mildenberger and Tingley 2017; Fong and Grimmer 2016). Mixed-membership models have also been used to describe political relationships using the relative emphasis of different topics in text as the dependent variable. For example, Bagozzi and Berliner (2016) examines attention to different issues vary over time in State Department Reports Genovese (2017) investigate the relationship between statements made by businesses and governments regarding climate change and sectoral levels of pollution and trade.

Each document can be represented as a vector of topic proportions, θ_d. For example, in a model of the Environmental Protection Agency's budget justifications, "sustainable," "climate," "carbon," and a dozen other words may co-occur in the same documents and be assigned high probabilities of being in a topic, which may be interpreted as a topic about climate change. Each document, representing an agency or subcommittee in a single year, has a proportion of words with a high probability of belonging to what we may call the *climate change* topic (if such a topic emerges). This may be a relatively high portion for Environmental Protection Agency documents and a lower portion for the House Appropriations Subcommittee on the Environment, Interior, and Related Agencies.

These proportions vary in each document type from year to year, providing a way to capture what Jones and Baumgartner (2005) call *attention allocation*, the change weights on policy images and issues: in this case, what the Environmental Protection Agency ought to do.

The results below use the LDA approach to infer document topic proportion and word-in-topic probabilities. We observe the total number of unique words $(w_1, ..., w_W)$ in the vocabulary of all documents and $w_{i,d}$ is the word observed at the ith token in document d. All texts are "tokenized" by giving each word a unique index i. If token i belongs to topic k, then the probability that the token is word w is the topic-specific probability $\pi_{k,w}$. At the document level, $\theta_{k,d}$ is the estimated proportion of topic k for document d (Blei et al. 2003).

Figure 7 shows the year-to-year change in each θ_k for an arbitrary set of 45 LDA topic models with K set at 5 through 45. While this approach does not seem to show spikes in change of word-use in

scratch (Grimmer and King 2011) and improves the consistency of estimated covariate effects Roberts et al. (2014).

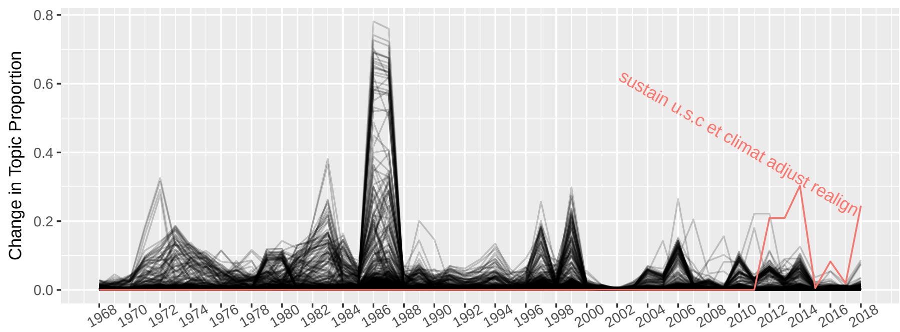
In my case, it is likely that budget reports for each agency and each committee are generated by a very similar process each year, making it better to assign each agency and committee a unique prior rather than assuming that all documents arise from the same distribution of words. The rate of use of each word in a topic is allowed to vary by the agency or committee who wrote it.

Figure 7: Measuring Similarity and Difference Over Time with Topic Models

Text:	$\boxed{\text{Year } t}$		$\boxed{\text{Year } t+1}$	Change
θ proportions of K topics:	$\theta_{1:k,t}$	$\longrightarrow$	$\theta_{1:k,t+1}$	$\lvert\theta_{1,t+1}-\theta_{1,t}\rvert,\ \lvert\theta_{2,t+1}-\theta_{2,t}\rvert,\ ...\lvert\theta_{k,t+1}-\theta_{k,t}\rvert$

presidential transition years, it does show a large spike during the Reagan administration when major shifts of EPA priorities led to the mass resignation of employees, probably the most tumultuous time in the agency's history. Additionally, the topic that changed most dramatically in the transition between the Obama and Trump administrations is one of the few topics for which the word "climate" is among the most frequent and exclusive terms identified by the FREX algorithm (Roberts et al. 2014).

Figure 8: Year-to-year Change in Environmental Protection Agency Budget Justification for 45 LDA Topic Models (K = 5:50), Fiscal Years 1967-2018



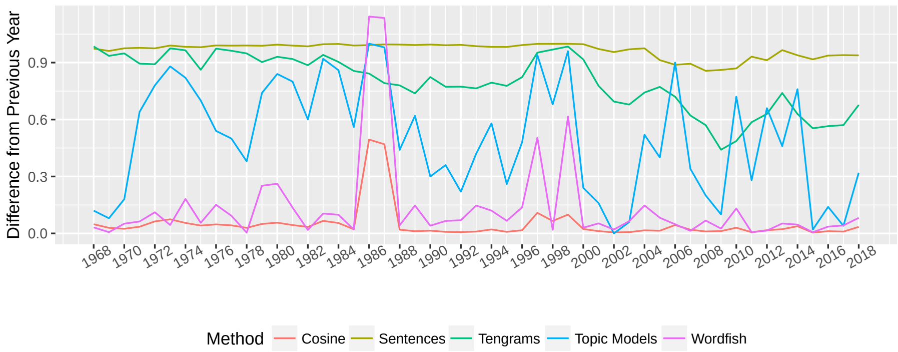
3.4 Contrasting Methods

3.4.1 Attention Allocation

None of the off-the-shelf methods to measure document similarity and difference performed well with respect to baseline expectations of capturing partisan differences. Results for each of the 71 agencies and 4 subcommittees are as uninformative as the results for the EPA shown in figure 9.

All three term frequency based methods (Cosine Similarity, Topic Models, and *Wordfish*) capture the discontinuity in the Reagan administration, but otherwise do not align with presidential transitions. The ten-gram matching consistently captures more text reuse than sentence matching. It finds that 30-50% of EPA budget justifications remain the same in the last 10 years, likely an underestimate given the requirement for exact matching. Both matching methods show increased reuse over time, likely reflecting improved text quality, with fewer errors due to text recognition of applied to paper documents.

Figure 9: Year-to-year Change in Environmental Protection Agency Budget Justification



3.4.2 Agenda Setting

Testing hypotheses about agenda setting will require better methods of estimating document similarity and change. Nevertheless, results based on the approaches described above are presented for illustrative purposes.

Rather than comparing year-to-year versions of each document, documents are here compared to the document that most plausibly influenced it.

Normalized cosine similarity and *Wordfish* distance are now calculated between pairs of documents from each agency and congressional committee in the same year and between each committee and the following year's agency budget justification. Figure 11 focuses on similarity rather than difference, so dimensional scaling similarity are calculated as 1 - *normalized Wordfish score*.

Topic proportions are now the average absolute value of the difference between the topic proportions in pairs of documents from each agency and congressional committee in the same year and between each committee and the following year's agency budget justification. To create a measure of similarity, this is subtracted from 1, for example, Mean$(1 - |\theta_{k,t,House} - \theta_{k,t,Agency}|)$ and Mean$(1 - |\theta_{k,t+1,Agency} - \theta_{k,t,House}|)$.

Figure 10 illustrates how such an approach would be used to test Hypotheses 1 and 2 to be tested with a time series model. Figure 11 plots average similarities in topic proportions alongside other measures of simility: cosine similarity, *Wordfish* scale similarity, and the percent of words in copied sentences and 10-grams.

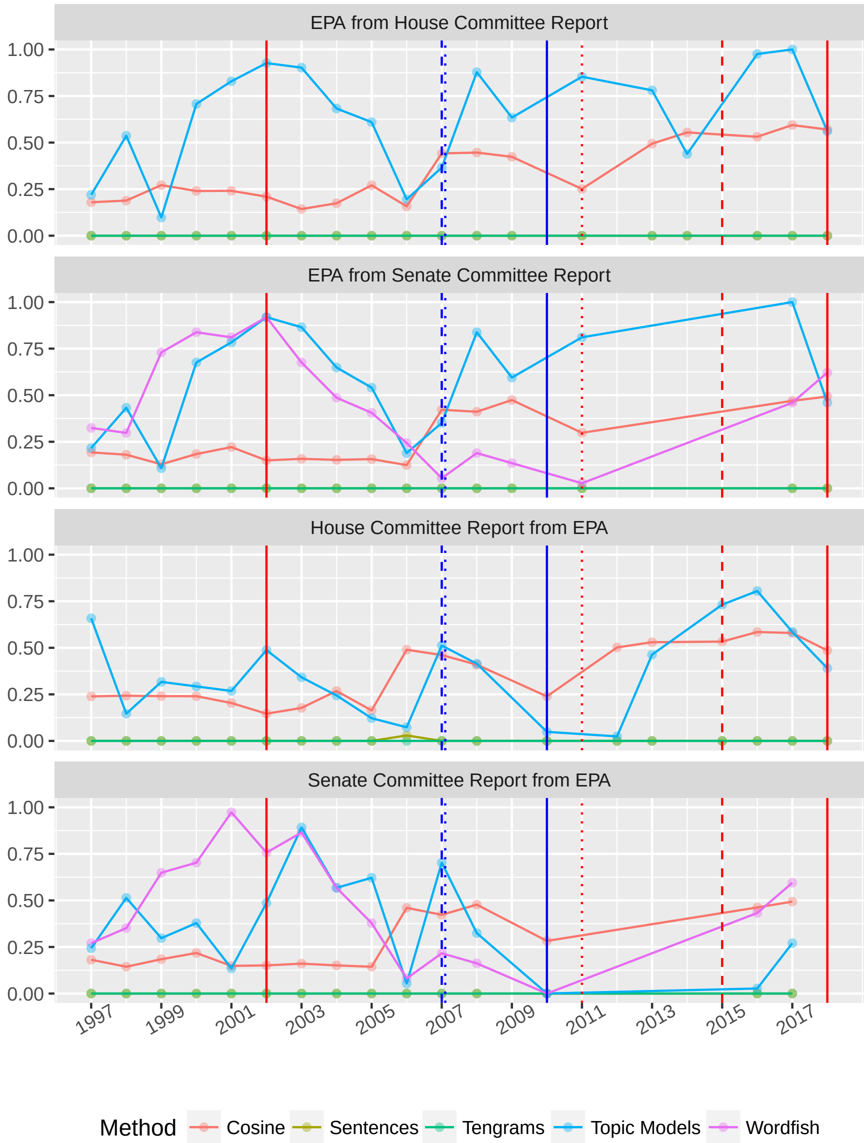
Figure 10 shows that very little text is copied directly across venues, but all three term-frequency

Figure 10: Measuring Agenda Setting with Topic Models

Text:	$\boxed{\text{President's Budget}}$	$\xrightarrow[H1]{}$	$\boxed{\text{Appropriations Report}}$
θ proportions of K topics:	$\theta_{1:k,t,Agency}$		$\theta_{1:k,t,Senate}$ $\theta_{1:k,t,House}$
	$\boxed{\text{Appropriations Reports}}$	$\xrightarrow[H2]{}$	$\boxed{\text{Next Year's President's Budget}}$
	$\theta_{1:k,t,Senate}$ $\theta_{1:k,t,House}$		$\theta_{1:k,t+1,Agency}$

based measures indicates indicated substantial variation in the similarity of word use between agencies and Congress. Unsuprisingly, given the poor relationship between these measures and partisan change, these measures do not exhibit clear patterns in cross-venue relationships.

Figure 11: Assessing the Relationship Between Congress and the EPA Over Time



4 Next Steps

Next steps include combining word-frequency and word-sequence information into the same model and less arbitrary selection of topics.

4.1 Topic Selection

While all topic models contain information, some models contain more of the information about relevant policy differences. Rather than simply averaging the change in topic proportion across and arbitrary number of topic models, selecting or weighting more informative models may reduce noise.

Given the arbitrariness of the number of topics selected and the potential instability in estimation, one key validation step is to demonstrate that topics make sense (Grimmer 2013). There is no guarantee that an unsupervised model will identify meaningful topics. As noted modeling strategies can help but are still no guarantee that meaningful latent topics will be recovered.

When topic models are used to generate independent variables, interpretation of topic content naturally receives attention. However, if scholarship moves more in the direction of hypothesis testing where theories focus on the similarity, difference, or change in topic distribution between documents, there is a risk that topic content could become hidden behind correlations of topic proportion means. To avoid this, two validations strategies have been suggested: (1) demonstrate that different algorithms produce similar topics and (2) establish that variations in topic emphasis across time or venues correlate with real-world events (Grimmer and King 2011; Roberts et al. 2014; Blei and Lafferty 2009; Quinn et al. 2010; Wilkerson and Casas 2016). Quinn et al. (2010) validate their classification of Senate speeches against previous findings from hand-coding approaches. Wilkerson and Casas (2016) suggest that, because topic selection is arbitrary, reporting and validating a single model is insufficient. They argue that topic selection and validation should explicitly address the robustness of topics across specifications. Beyond validation, they illustrate that focusing on topic robustness can help interpret results by grouping potential topics into robust metatopics. Thus, in addition to interpretability and face validity, reliability across model specifications is an important standard for unsupervised approaches.

Once the modeling approach has been improved (see next subsection) identifying the more informative and interpretable topic models is an essential next step in this project.

Topic-specific estimates of change over time and potential agenda setting between Congress and agencies may allow for issue specific predictions for the impact of a presidential transition or change in committee control. Partisanship and liberal-conservative ideology scores are often used to predict issue support. Scholars have coded agencies as liberal, conservative, or neutral as well as more nuanced scores based on surveys (Clinton and Lewis 2008). Benoit and Herzog (2015) also estimate ideological positions of lawmakers using transcripts of budget debates, which could be extended to include budget justification texts and testimony. Topic modeling may allow for more targeted approach. After estimating the major topics of discussion for each agency's budget, I will code each party and subcommittee chair as in favor (+1), neutral (0), or opposed (-1) to spending on each issue by triangulating available information including new stories, press releases, and interest group scorecards, for example by searching the name of the member and the most frequent and unique exclusive words representing the topic. This will create issue-specific predicted support, allowing topic-specific tests of agenda setting hypotheses.

4.2 Combining Methods Based on Text Reuse and Word Frequency

Second, if text reuse can provide a reliable measure of similarity and difference across budget texts, this information may be incorporated into topic models to improve the estimation of meaningful topics and topic distributions. Previous work has included document attributes such as the author, year, or treatment condition (Roberts et al. 2014), document network structure such as citations (Chang et al. 2009), word correlations (Blei and Lafferty 2005), sentiment of words, constraints based on intuitions ("lexical priors") on the distribution of words over topic (Jagarlamudi et al. 2012), or combinations of these kinds of information (Kang et al. 2014).

The extension of LDA that Chang et al. (2009) call a Relational Topic Model may be especially well suited for incorporating document-level information about text reuse. This model includes a binary random variable for each document pair that is conditioned by the latent space that also determines topic proportions. Thus, the document generating process has an extra step:

1. Draw topic proportions $\theta_d|\alpha \sim Dir(\alpha)$.
2. For each word $w_{d,n}$:
   (a) Draw assignment $z_{d,n}|\theta_d \sim Mult(\theta_d)$
   (b) Draw word $w_{d,n}|z_{d,n,\beta_{1:K}} \sim Mult(\beta_{z_{d,n}})$.
3. For each pair of documents d, d':
   $y|z_d, z_{d'} \sim \psi(\cdot|x_d, z_{d'})$.

Chang et al. (2009) demonstrate this model with respect to academic documents and citations, but political influence may be modeled similarly. For example, documents may be linked by one citing the other or plagiarizing the other.

In addition to document-level covariates used by the Relational Topic Model and Structural Topic Model, text reuse methods provide token-level information that may be used in the likelihood function. Acree et al. (2016) suggest that if cosine similarity is to be used to estimate document similarity, it should be weighted with local alignment, incorporating local alignment into topic modeling strategies could accomplish a similar aim with more interpretable results.

5 Conclusion

Policy texts reflect the "quality" aspects of what government officials do Grimmer (2013) and qualitative policy content is also a key policy output (Mansbridge 2003). Because words give meaning to political ideas, associated texts give meaning to line items and votes. In the case of budget justifications, this meaning is especially interesting because it can be exactly matched with numeric budget allocations.

The off-the-shelf methods for assessing change over time and relationships between documents do not appear to work well for the case of budgeting, despite significant amounts of well-structured texts.

Combining classification algorithms like topic models with matching methods that identify text reuse may have broad potential in political science. Methods based on word frequency and text reuse capture different but potentially relevant aspects of the relationships across policy texts that may help identify what is flying under the radar, what is copied from elsewhere or otherwise receiving special attention, and ultimately who is driving the substance of policy change.

References

Acree, Brice, Joshua Jansa, and Kelsey Shoub. 2016. "Comparing and Evaluating Cosine Similarity Scores, Weighted Cosine Similarity Scores, and Substring Matching."

Adler, E. Scott, and John D. Wilkerson. 2012. "Congress and the Politics of Problem Solving." In *Part 1,* 3–18.

Adler, ES, and JS Lapinski. 1997. "Demand-side theory and congressional committee composition: A constituency characteristics approach." *American Journal of Political Science* 41 (3): 895–918.

Bagozzi, Benjamin E., and Daniel Berliner. 2016. "The Politics of Scrutiny in Human Rights Monitoring: Evidence from Structural Topic Models of US State Department Human Rights Reports." *Political Science Research and Methods* (October): 1–17.

Baumgartner, F R, and B D Jones. 1991. "Agenda Dynamics and Policy Subsystems." *Journal of Politics* 53 (4): 1044–1074.

Bendor, Jonathan, Amihai Glazer, and Thomas Hammond. 2001. "Theories of delegation." *Annual review of political science* 4 (1): 235–269.

Benoit, Kenneth, and Alexander Herzog. 2015. "Text Analysis: Estimating Policy Preferences From Written and Spoken Words."

Benoit, Kenneth, Kohei Watanabe, Paul Nulty, Adam Obeng, Haiyan Wang, Benjamin Lauderdale, and Will Lowe. 2017. *quanteda: Quantitative Analysis of Textual Data.*

Berry, Christopher R, Barry C. Burden, and William G Howell. 2010. "The President and the Distribution of Federal Spending." *American Political Science Review* 104 (04): 783–799.

Blei, D, AY Ng, and MI Jordan. 2003. "Latent dirichlet allocation." *Journal of Machine Learning Research* 3 (1): 993–1022.

Blei, David M, and John D Lafferty. 2005. "Correlated Topic Models." In *Advances in Neural Information Processing Systems 18 (NIPS).*

———. 2009. "Topic Models." *Text mining: classification, clustering, and applications* 10 (71): 34.

Bolton, Alex, and Sharece Thrower. 2015. "The Constraining Power of the Purse: Executive Discretion and Legislative Appropriations *."

Brady, Jacob, Robert Neihesel, and Kevin Richard Stout. 2016. *Stimulating Presidential Support: The American Recovery and Reinvestment Act, Presidential Pork, and Vote-buying in Congress.* Chicago, IL.

Brookhart, Jennifer L, and Alexander Tahk. 2015. "The Origin of Ideas."

Butler, Daniel M., and Eleanor Neff Powell. 2014. "Understanding the Party Brand: Experimental Evidence on the Role of Valence." *The Journal of Politics* 76 (02): 492–505.

Carpenter, Daniel P. 2001. *The forging of bureaucratic autonomy: Reputations, networks, and policy innovation in executive agencies, 1862-1928.* Princeton University Press.

Carpenter, Daniel. 2014. *Reputation and power: organizational image and pharmaceutical regulation at the FDA.* Princeton University Press.

Casas, Andreu, Matthew J Denny, and John Wilkerson. 2017. "Legislative Hitchhikers : Re-envisioning Legislative Productivity and Bill Sponsorship Success."

Chang, Jonathan, Jordan Boyd-Graber, Chong Wang, Sean Gerrish, and David M Blei. 2009. "Reading Tea Leaves: How Humans Interpret Topic Models." In *Neural Information Processing Systems,* 288–96. Cambridge, MA: MIT Press.

Clinton, Joshua D., and David E. Lewis. 2008. "Expert Opinion, Agency Characteristics, and Agency Preferences." *Political Analysis* 16, no. 01 (January): 3–20.

Cox, Gary W., and Mathew Daniel McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives.* 336. Cambridge [England] ; New York, N.Y.: Cambridge University Press.

Curry, James M. 2015. *Legislating in the dark : information and power in the House of Representatives.* 274.

Denny, Matthew J., and Arthur Spirling. 2017. "Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It."

Fisher, Louis. 2015. *Presidential Spending Power.* Princeton, NJ: Princeton University Press.

Fong, Christian, and Justin Grimmer. 2016. "Discovery of Treatments from Text Corpora": 1–20.

Garrett, Kristin N, and Joshua M Jansa. 2015. "Interest Group Influence in Policy Diffusion Networks." *State Politics and Policy Quarterly:* 1–31.

Genovese, Federica. 2017. "Sectors, Pollution, and Trade: How Industrial Interests Shape Domestic Positions on Global Climate Agreements I am thankful for comments and feedback from."

Grimmer, Justin. 2013. "Appropriators not position takers: The distorting effects of electoral incentives on congressional representation." *American Journal of Political Science* 57 (3): 624–642.

Grimmer, Justin, and Gary King. 2011. "General purpose computer-assisted clustering and conceptualization." *Proceedings of the National Academy of Sciences of the United States of America* 108 (7): 2643–2650.

Hertel-Fernandez, Alexander, and Konstantin Kashin. n.d. "Capturing Business Power Across the States with Text Reuse."

Jagarlamudi, Jagadeesh, Hal Daumé Iii, and Raghavendra Udupa. 2012. "Incorporating Lexical Priors into Topic Models." In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics,* 204–213. Avignon, France.

Jones, Bryan D, and Frank R Baumgartner. 2005. *The Politics of Attention: How Government Prioritizes Problems.* Chicago, IL: University of Chicago Press.

Kang, Dongyeop, Youngja Park, and Suresh N. Chari. 2014. "Hetero-Labeled LDA: A Partially Supervised Topic Model with Heterogeneous Labels." In *Machine Learning and Knowledge Discovery in Database. ECML PKDD 2014. Lecture Notes in Computer Science, vol 8724.* Edited by T. Calders, F. Esposito, E. Hüllermeier, and R. Meo, 640–655. Berlin, Heidelberg: Springer.

Kingdon, John W. 1995. *Agendas, alternatives, and public policies.* 2nd. xiv, 254. New York: HarperCollins College Publishers.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97 (2): 311–331.

Lee, Frances E. 2016. *Insecure majorities : Congress and the perpetual campaign.* 266.

———. 2000. "Senate Representation and Coalition Building in Distributive Politics." *American Political Science Review* 94 (1): 59–72.

Linder, Fridolin, Bruce Desmarais, Matthew Burgess, and Eugenia Giraudy. 2017. "Text as Policy: Measuring Policy Similarity through Bill Text Reuse."

Lowi, Theodore. 1967. "The Public Philosophy: Interest-Group Liberalism." *The American Political Science Review* 61 (1): 5–24.

Mansbridge, Jane. 2003. "Rethinking Representation." *American Political Science Review* 97 (4).

McCubbins, Mathew D, and Thomas Schwartz. 1984. "Congressional oversight overlooked: police patrols versus fire Alarms." *American Journal of Political Science* 28 (1).

McCubbins, Mathew D, Roger G Noll, and Barry R Weingast. 1987. "Administrative procedures as instruments of political control." *Journal of Law, Economics, & Organization* 3 (2): 243–277.

Mildenberger, Matto, and Dustin Tingley. 2017. "Beliefs about Climate Beliefs: Second-Order Opinions in the Climate Domain." *British Journal of Political Science.*

OMB. 2016. *Circular No. A-11 Revised.*

Powell, Eleanor Neff, and Justin Grimmer. 2016. "Money in Exile: Campaign Contributions and Committee Access." *The Journal of Politics* 78 (4): 974–988.

Quinn, Kevin M, Burt L Monroe, Michael Colaresi, Michael H Crespin, and Dragomir R Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54 (1): 209–228.

Roberts, Margaret E, Brandon M Stewart, and Tingley Dustin. 2017. "stm: R Package for Structural Topic Models." *Journal of Statistical Software* VV (II).

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–1082.

Shepsle, Kenneth A., and Barry R. Weingast. 1987. "The Institutional Foundations of Committee Power." *American Political Science Review* 81 (1): 85–104.

Slapin, Jonathan B, and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52 (3): 705–722.

Smith, Temple F., and Michael S. Waterman. 1981. "Identification of common molecular subsequences." *Journal of Molecular Biology* 147, no. 1 (March): 195–197.

Whittington, Keith E, and Daniel P Carpenter. 2003. "Executive power in American institutional development." *Perspectives on Politics* 1 (03): 495–513.

Wildavsky, Aaron B. 1964. *Politics of the budgetary process.* Boston, MA: Little Brown.

Wilkerson, John D., and Andreu Casas. 2016. "Large-scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science:* 1–18.

Wilkerson, John, David Smith, and Nicholas Stramp. 2015. "Tracing the Flow of Policy Ideas in Legislatures: A Text Reuse Approach." *American Journal of Political Science* 59 (4): 943–956.

Yackee, J W, and S W Yackee. 2009. "Divided government and US federal rulemaking" [in English]. *Regulation & Governance* 3 (2): 128–144.

Appendix

Figure 12: Estimated Impact of Preprocessing Steps on Topic Model Results using PreText()