An Automated Method of Topic-Coding Legislative Speech Over Time with Application to the 105th-108th U.S. Senate*

Kevin M. Quinn†  Burt L. Monroe‡  Michael Colaresi§  Michael H. Crespin¶  Dragomir R. Radev∥

July 18, 2006

Abstract

We describe a method for statistical learning from speech documents that we apply to the Congressional Record in order to gain new insight into the dynamics of the political agenda. Prior efforts to evaluate the attention of elected representatives across topic areas have largely been expensive manual coding exercises and are generally circumscribed along one or more features of detail: limited time periods, high levels of temporal aggregation, and coarse topical categories. Conversely, the Congressional Record has scarcely been used for such analyses, largely because it contains too much information to absorb manually. We describe here a method for inferring, through the patterns of word choice in each speech and the dynamics of word choice patterns across time, (a) what the topics of speeches are, and (b) the probability that attention will be paid to any given topic or set of topics over time. We use the model to examine the agenda in the United States Senate from 1997-2004, based on a new database of over 70 thousand speech documents containing over 70 million words. We estimate the model for 42 topics and provide evidence that we can reveal speech topics that are both distinctive and inter-related in substantively meaningful ways. We demonstrate further that the dynamics our model gives us leverage into important questions about the dynamics of the political agenda.

* An earlier version of this paper was presented to the Midwest Political Science Association and was awarded the 2006 Harold Gosnell Prize for Excellence in Political Methodology. We would like to thank Steven Abney, Scott Adler, Scott Ainsworth, Frank Baumgartner, Ken Bickers, Barry Burden, Will Howell, Glen Krutz, Frances Lee, Iain McLean, Nate Monroe, Steven Purpura, Gisela Sin, and seminar participants at the Universities of Rochester and Michigan for their comments on earlier versions of the paper. We would like to give special thanks to Cheryl Monroe for her crucial volunteer contributions toward development of the Congressional corpus in specific and our data collection procedures in general. We would also like to thank Jacob Balazer (Michigan), Tony Fader (Michigan), T. J. Knezek (Michigan State), Jen Miller (Michigan) for research assistance and Amber Boydstun (Penn State) for her assistance in adapting data from the Policy Agendas Project to our purposes. This paper is based upon work supported by the National Science Foundation under Grant No. 0527513. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

†Department of Government and Institute for Quantitative Social Science, Harvard University, kevin.quinn@harvard.edu.
‡Department of Political Science, The Pennsylvania State University, and Center for Political Studies, University of Michigan, monroeb@umich.edu.
§Department of Political Science, Michigan State University, colaresi@msu.edu.
¶APSA Congressional Fellow and Department of Political Science, University of Georgia, crespin@msu.edu.
∥School of Information and Department of Electrical Engineering and Computer Science, University of Michigan, radev@umich.edu.
1 Introduction

The verbatim records of democratic legislatures represent a source of untapped information of singular importance for the study of democratic societies. In legislative records, we observe – as we can nowhere else – a slowly changing set of political elites defining and discussing the issues of the day. Legislative speech records tell us what day, and in some cases what time of day, these discussions occurred and we have such verbatim records for decades or even, in a few nations, centuries. No other source – not election returns, not opinion polling, not newspapers, not legislation, not party manifestos – provides such a potentially rich and consistent source of information about the substantive dynamics of democracy.

Despite this, legislative records have remained largely untapped as a data source. This appears to be, perversely, because they contain too much information. A typical year of any legislative record includes tens of thousands of speeches, and tens of millions of spoken words. It is more or less impossible for a single individual to read, much less absorb or analyze in any way, the entire record of a given legislature as quickly as it is produced. Attempts to analyze The Congressional Record, for example, are generally limited to small segments of time or very small samples, often requiring extensive and expensive manual coding efforts\(^1\). Increasingly, however, legislative records are available in electronic form. With advances in statistical learning and computational linguistics (as well as computational hardware), it now becomes possible to put such records to broad scientific use.

Our focus in this paper is to leverage the temporal richness of these legislative speech data to illuminate the dynamics of the political agenda. Specifically, we introduce a new statistical method for determining what topics occupy the attention of legislators and when. We apply the method to a new and massive dataset based on patterns of word choice in speech recorded in the Congressional

---

\(^1\)For example, Hill and Hurley (2002) use two manual coders to place a sample of 2,000 Senate speeches from 1990 into three categories. Box-Steffensmeier et al. (1997), in a study that examines timing of position-taking announcements (including floor speech) on NAFTA during the 103rd Congress note that “Data on timing, to be adequate to the task, are collected at substantial cost” (p. 330).
Record and we demonstrate how this can be used to gain new and powerful leverage on a variety of longstanding questions in political science. How do elected leaders distribute their attention (Jones and Baumgartner, 2004)? What are the dynamics of this process? Is policy making incremental, or explosive, or both, characterized by punctuations (Wlezien, 1995; Baumgartner and Jones, 1993; Jones and Baumgartner, 2005)? How are policy areas interrelated (Katznelson and Lapinski, 2006)? How do surprises in the agenda reverberate across different topic areas? What are the democratic properties of the agenda – under what circumstances do leaders push or follow public attention (Jacobs and Shapiro, 2000; Stimson et al., 1995)?

To answer such questions, we must have a measure of how the attention of legislators is distributed across policy topics, political issues, and other matters that compete for their scarce time. As a source for such measures, legislative speech differs from the extant alternatives in a number of important ways. If we are interested in any dynamic processes that might occur between elections, election-based measures – voter perceptions from surveys like the National Election Studies; party declarations in manifestos or platforms (Budge et al., 2001) – will be insufficient. Alternatively, one might look to the in-session activities of legislators, such as speeches, hearings, bills, and votes.

In the study of American politics, the state of the art on agenda measurement is the Policy Agendas Project (Jones et al., N.D.) [, hereafter PAP], discussed in a long-running series of publications (most prominently Baumgartner and Jones (1993), Baumgartner and Jones (2002), and Jones and Baumgartner (2005)). They measure the policy agenda through a massive manual coding effort on Congressional hearings. The results are, for most purposes, annually aggregated indicators of total (counts) and relative attention (% of counts) given through hearings to 19 (or

2 Manifestos have an additional feature, problematic for the sorts of analyses we do here. Parties are compelled to declare their policies in every issue area (the Beer-Drinker’s Party will describe its foreign policy), even though they would and will focus on many fewer priority areas in practice. This makes it more difficult to evaluate differences in priorities than does inter-election behavior.

3 They also measure the “public agenda” through the Gallup organization’s “Most Important Problem” question, the “systemic agenda” through coding a sample of editorials from The New York Times, and policy itself through legislation and budgets.

4 Some indicators have a month or date stamp, but there is little or no variation across months, let alone weeks or days. Most of the PAP research is based on annual measures and addresses questions of long-term dynamics. Our approach is complementary: subannual measures allowing studies of short-term dynamics.
more for some purposes) major topic areas, further subdivided into over 200 minor topics. The closely related Congressional Bills Project (Adler and Wilkerson, 2006), [hereafter CBP], provides a census of House and Senate bills which have been manually coded according to the PAP policy codes and could easily be incorporated into similar analyses; Purpura and Hillard (2006) have developed methods for automated coding of bills according to the PAP scheme. Rollcall votes have been manually coded by a variety of other coding schemes (e.g. Poole and Rosenthal, 1997; Rohde, 2004; Lee, 2006).

The speech-based measures we introduce here differ from hearings, bills, and votes in a variety of important ways. First, speech is observable to the day (when the legislature is in session), and is voluminous and variable enough to allow significantly greater leverage on subannual dynamics. For example, consider the issue of abortion in the U.S. Senate from 1997-2004, the period we examine below. The PAP data show three Senate hearings, in total, on abortion over the eight years. The CBP data end in 1998, but show nine abortion bills in 1997-98; a rough review suggests approximately 30 abortion-related bills over the entire period. A generous definition of “on abortion” shows 21 rollcall votes on abortion during the period. By contrast, our speech data show over 500 abortion speeches, containing just under one million words.\(^5\)

Second, speech is observable to the individual, and is more voluminous and variable in cross-section. We can observe wide variation in the patterns of speech among individual legislators, and arbitrary groupings thereof. We can observe not just the macrodynamics of the legislature as a whole, but also the dynamics of parties, individuals, or any other subaggregate of interest. For a vote, we can observe for individuals only yes or no (in addition to abstentions, absences, and the like). For a bill, we can observe only sponsorship, cosponsorship, or not. Hearings are generally called by committee chairs, and so can only be attributed to a small number of individuals, all in the majority.

\(^5\)Schonhardt-Bailey (2006) shows roughly this number of speeches on abortion in just a few days in 2003 alone. The difference is primarily in the operationalization of “speech”, as she defines a new speech each time a speaker is interrupted and we do not, and partly in the classification of speeches as “on abortion”.
Third, the fact that speeches are composed of words provides them with both substantive content and a basis for comparison with one another. Put simply, speeches try to tell you what they are about. This is also partially true of hearings, but it is less true of bills and largely false in the case of rollcall votes. For some pieces of legislation, the most meaningful votes in Congress are procedural votes – on special rules, for example – that have no apparent policy meaning in their language whatsoever.\footnote{Conversely, the debate surrounding such votes, often will discuss the affected substance.} Moreover, since speeches are compositions of identifiable repeated elements (words from a vocabulary), we can determine how they are related to one another topically and develop consistent notions of dynamic change. The lack of these features in rollcalls – each vote, each bill, is \textit{sui generis} – is reflected in the difficulty rollcall analysts have with naming the dimensions recovered from ideal point estimation procedures (Jackman, 2001; Poole, 2005) and with conveying consistent notions of dynamics (Martin and Quinn, 2002; Poole, 2005). If this is a bit abstract at this juncture, the power of the self-defining nature of speech becomes quite clear in our examples below.

Fourth, the decision to make a speech on a particular topic is subject to very different strategic calculations and selection mechanisms, particularly in contrast to votes. In the case of rollcall votes, the work with NOMINATE and successors (Poole and Rosenthal, 1991, 1997; Poole, 2005) has found that current legislative voting behavior can be described almost in its entirety by one increasingly partisan ideological dimension. But it is widely believed that this is, at best, the end result of a complex process by which high-dimensionality policy options and preferences are mapped into the dominant ideologies (Poole, 2005; Potoski and Talbert, 2000; Ringe, 2005), and, at worse, an artifact of strategic behavior (Kim, 2006; Clinton and Meirowitz, 2003; Rohde, 2004; Sinclair, 1995). On the latter point, for example, if majorities exercise agenda gate-keeping powers as described by Cox and McCubbins (2005), and can select only those rollcall subjects that will command a “majority of the majority” (now entrenched in the House as the so-called “Hastert rule”), the results will appear unidimensional (and polarized by partisanship) regardless of the true
dimensionality of the space. Speech is not subject to the same harsh selection mechanisms, and is more revealing of the full dimensionality and complexity of the issues and preferences under consideration, a point already made by other legislative scholars (Schonhardt-Bailey, 2006; Ringe, 2005). We note in particular, that when majorities control access to hearings and the voting agenda, speech can still better illuminate the priorities and preferences of minorities and oppositions.

But does speech contain useful information in the first place? We have been accused on occasion of taking speech too seriously. Generally, for a message to convey information, it must be costly to send. Otherwise, in equilibrium, the message is a meaningless babble. Surely, this argument goes, the emptiness of the debating chamber, the obscurity of C-SPAN drama, and the sheer volume we have noted, argue for this interpretation. But speech is costly in both supply and demand. The total amount of floor time is limited and, at least on occasion, enviously sought. Legislators shoulder considerable opportunity costs when they choose to speak rather than pursue some other activity (Hall, 1996; Wawro, 2000). Legislative speech is rarely for the audience in the chamber, but choice excerpts appear in almost every evening newscast. Legislators frequently quote their own floor speeches in things like constituency newsletters, demonstrating the attention they are giving to particular issues. Poorly considered speech can be used later by opponents (e.g., “I voted for the 87 billion dollars before I voted against it”). So legislative speech is a costly signal, and as such should be expected to contain information the speaker wants to convey. In any case, an assertion of the random meaninglessness of legislative speech can be taken as a clear empirical hypothesis, one that is dramatically falsified in our results here.

We have also been accused of not taking speech seriously enough. Cutting up a speech into words, as we do here, without any concern for word order much less its persuasive features as a piece of argumentative rhetoric, does grave violence to language and meaning. That is, the only way to extract meaning from speech is through careful manual analysis (as in Mucciaroni and Quirk

---

7 This varies by legislature as well. Time in the 650+ member of the House of Commons is extremely highly valued, for example.

8 We note further that a sufficiently large political science audience – that is, an audience of two or more – will generate both of these criticisms as a matter of course.
(2006)), or further, through close exegesis or hermeneutics. But this is a false choice, analogous to saying one may examine the weather satellite map or look out the window, but not both. Like weather satellites, our approach accepts some level of local error, while providing an accurate big picture, an understanding of larger forces at play, an ability to forecast, and an ability to recognize when an event is surprising. No one would argue not to look out your window when deciding whether to carry an umbrella, nor would anyone argue for turning off the satellites. And again, the assertion that our approach cannot extract any meaning constitutes an empirical hypothesis that is clearly falsified by our results below.

Our approach to extracting topic information from speech differs from more familiar methods of content analysis in two fundamental ways: it is unsupervised and it is dynamic. Put abstractly, there are four basic approaches to categorizing texts, distinguished by the relative level of human versus automated processes; manual (concept-based) coding, dictionary-based coding, supervised learning, and unsupervised learning. In manual coding, the analyst develops a set of conceptual categories and human coders label each text accordingly, with examples including PAP, CBP, Rohde (2004); Lee (2006). In dictionary-based coding, analysts use a manual coding step for a subset of documents to develop a dictionary that maps terms into categories. The dictionary is then applied computationally to the texts to be coded. Political science examples include Laver and Garry (2000); Gerner et al. (1994). Under supervised learning, a subset of texts is manually coded and then treated as a training set. A statistical algorithm evaluates what features of the training texts best identify categories and evaluates new texts according to these features. Spam filters are a well-known example, attempting to infer whether a given text is “spam” or “not-spam” on the basis of examples of each. A recent political science example is Purpura and Hillard (2006). Unsupervised learning, which is our approach here, has no predefined set of categories, dictionaries, or training data. The goal is to place texts statistically into categories that maximize internal coherence. This is considerably less expensive, as there is no manual step, except perhaps in the ex-post evaluation and interpretation of the found categories. This has the advantage of letting the data speak for
themselves with regard to natural categorization and the disadvantage of not necessarily capturing particular distinctions of interest to the analyst (e.g., spam / not-spam).

As a sidebar, we should also note the contrast between our task of categorization and that of information retrieval. The most common forms of information retrieval involve either a manually-defined index system or an automated keyword search of a database. One might, for example, measure attention to health policy by counting uses of the word “health” in a database, or lookup documents already indexed by the keyword “health.” This is a perfectly reasonable approach for some problems.\(^9\) Baumgartner and Jones (1993); Adler and Wilkerson (2006), however, provide compelling summaries of the problems with this approach for many problems including agenda research, with particular emphasis on how the rules governing such systems shift and drift arbitrarily over time. We would emphasize further that categorization places each text into one of a set of mutually exclusive and exhaustive categories, allowing for an exact tracking of how attention is allocated across categories. With information retrieval, texts without the selected search or index terms will never appear and those with more than one will appear more than once. The types of analyses we conduct in this paper, like those in the Baumgartner and Jones work, are not possible with information retrieval.

And, of course, many hybrids of these approaches are possible. Schonhardt-Bailey (2006), for example, uses unsupervised correspondence analysis and clustering algorithms from the ALCESTE software to analyze the 2003 partial birth abortion debate. But, a necessary prior step for the analysis is identifying the speeches that are part of the particular debate. This is done through information retrieval – including only and all speeches given “Partial Birth Abortion Ban Act of 2003” subject headings in the Congressional Record – with some manual modifications – e.g., including speeches under the heading “The Right to Choose” and “Roe v. Wade”. This approach not only misses the latter if not caught manually, but misses anything not given a focused subject head (like speeches accompanying the bill’s introduction on February 14, labelled “Statements

\(^9\)For example, Maltzman and Sigelman (1996) use an electronic indexing system to estimate the number of speeches of one particular type during the 103rd Congress.
on Introduced Bills and Resolutions”); it also includes anything under the subject head, even if irrelevant (like speeches our approach below labels as procedural).\(^{10}\)

We should also note the contrast between categorization and procedures like factor analysis or scaling. The primary focus in rolcall work has been the measurement of legislator “ideal points”, or preferred locations in a policy or ideological space, and the identification of the dimensions of that space. The Comparative Manifestos Project has a similar goal with regard to the placement of political parties in a space. There are similar efforts underway with text (Laver et al., 2003; Schonhardt-Bailey, 2006; Gabel and Huber, 2000; Martin and Vanberg, 2006). We have developed such methods as part of our larger project here (Monroe and Maeda, 2004), but this is a different inferential task. Below we discuss how these two tasks are related and give an example of how they might be combined.

Our approach differs from approaches familiar in computational linguistics primarily in its exploitation of the dynamics of the legislative setting. As in a typical clustering problem, we assume that each speech delivered on a given day has some probability of being “about” one of \(K\) topics. We further assume that the distribution of attention is a dynamic process, related from one day to the next. We seek to determine, through the patterns of word choice in each document and the dynamics of word choice patterns across time, (a) what the topics of speeches are, and (b) the probability that attention will be paid to any given topic or set of topics over time. Our model differs from models applied to similar problems in computational linguistics – such as the latent Dirichlet allocation model of Blei et al. (2003), used for a variety of classification problems including the identification of topics in a set of scientific abstracts – on several dimensions, but varies most notably in the use of dynamics.

The remainder of the paper is organized as follows. In Section 2 we develop a statistical model for legislative speech data. Of particular interest is the dynamic hierarchical prior distribution for the

\(^{10}\)This has consequences for the analysis as well. Schonhardt-Bailey’s Figure 3 shows the party leaders, Frist (76.0) and Reid (36.0) as outliers together on a procedural dimension. But this reflects only their need to speak more procedural text, not anything about the particular language they use or positions they take when discussing abortion as such.
documents’ latent topic indicators. This portion of the model specification allows us to investigate the dynamics of legislative speech using well-developed time series tools (West and Harrison, 1997). The remainder of the paper deals with the application of our model to the spoken record of the 105th-108th U.S. Senate. In Section 3 we discuss how we convert the raw source of the Congressional Record into processed textual data and further into the statistical data required for the model. We apply the model to these data in Section 4. We make the case that the topic-codings produced by the model agree with generally recognized issue areas and that the relationships between the issue areas are also congruent with a common-sense understanding of American politics. We go on to show how the model can yield new insights about the dynamics of issue attention. The final section concludes with our plans for the future.

2 A Model for Dynamic Multi-Topic Speech

The data generating process that motivates our model is the following. On each day that Congress is in session a legislator can make speeches. These speeches will be on one of a finite number $K$ of topics. The probability that a randomly chosen speech from a particular day will be on a particular topic is assumed to vary smoothly over time. At a very coarse level, a speech can be thought of as a vector containing the frequencies of words in some vocabulary. These vectors of word frequencies can be stacked together in a matrix whose number of rows is equal to the number of words in the vocabulary and whose number of columns is equal to the number of speeches. This matrix is our outcome variable. Our goal is to use the information in this matrix to make inferences about the topic probabilities and how they change over time as well as the topic membership of individual speeches.

We begin by laying out the necessary notation. Let $t = 1, \ldots, T$ index time; $d = 1, \ldots, D$ index speech documents; $k = 1, \ldots, K$ index possible topics that a document can be on; and $w = 1, \ldots, W$ index words in the vocabulary. For reasons that will be more clear later, we also introduce the function $s : \{1, \ldots, D\} \rightarrow \{1, \ldots, T\}$. $s(d)$ tells us the time period in which document $d$ was put
into the *Congressional Record*. In addition, let $\Delta^N$ denote the $N$-dimensional simplex.

### 2.1 The Sampling Density

The $d$th document $y_d$ is a $W$-vector of non-negative integers. The $w$th element of $y_d$, denoted $y_{dw}$, gives the number of times word $w$ was used in document $d$.

We consider two equivalent probability models for $y_d$. First, if $y_d$ is generated from topic $k$ we could assume that

$$y_{dw} \sim \text{Poisson}(\exp(\alpha_d + \beta_{kw})),$$

Here $\alpha_d$ is related to the overall length, in total number of words, of document $d$ while $\beta_{kw}$ captures the log odds of word $w$ relative to some baseline. Equivalently, we can condition on the total number $n_d$ of words in each document. This generates a multinomial distribution for $y_d$:

$$y_d \sim \text{Multinomial}(n_d, \theta_k)$$

Here $\theta_k \in \Delta^{W-1}$ is the vector of multinomial probabilities with typical element $\theta_{kw}$. In what follows, we condition on $n_d$ for all $d$ and use the multinomial sampling density. This allows us to fit the resulting model more quickly. For purposes of interpretation, we will at some points below make use of the transformation

$$\beta_k = \left( \log \left( \frac{\theta_{k1}}{\theta_{k1}} \right) - c \right), \left( \log \left( \frac{\theta_{k2}}{\theta_{k1}} \right) - c \right), \ldots, \left( \log \left( \frac{\theta_{kW}}{\theta_{k1}} \right) - c \right),$$

where $c = W^{-1} \sum_{w=1}^{W} \log \left( \frac{\theta_{kw}}{\theta_{k1}} \right)$.

If we let $\pi_{tk}$ denote the marginal probabilities that a randomly chosen document is generated from topic $k$ in time period $t$ we can write the sampling density for all of the observed documents as:

$$p(Y|\pi, \theta) \propto \prod_{d=1}^{D} \sum_{k=1}^{K} \pi_{s(d)k} \prod_{w=1}^{W} \theta_{k1}^{y_{dw}}$$

As will become apparent later, it will be useful to write this sampling density in terms of latent data $z_1, \ldots, z_D$. Here $z_d$ is a $K$-vector with element $z_{dk}$ equal to 1 if document $d$ was generated
from topic \( k \) and 0 otherwise. If we could observe \( z_1, \ldots, z_D \) we could write the sampling density above as:

\[
p(Y, Z | \pi, \theta) \propto \prod_{d=1}^{D} \prod_{k=1}^{K} \left( \pi_{s(d)k} \prod_{w=1}^{W} \theta_{kw}^{z_{dk}} \right)
\]

### 2.2 The Prior Specification

To complete a Bayesian specification of this model we need to determine prior distributions for \( \theta \) and \( \pi \). We assume a conjugate Dirichlet prior for \( \theta \). More specifically, we assume

\[
\theta_k \sim Dirichlet(\lambda_k) \quad k = 1, \ldots, K
\]

For the data analysis below we assume that \( \lambda_{kw} = 1.01 \) for all \( k \) and \( w \).

The prior for \( \pi \) is more complicated. Let \( \pi_t \in \Delta^{K-1} \) denote the vector of topic probabilities at time \( t \). The model assumes that \( a \ priori \)

\[
z_{dt} \sim Multinomial(1, \pi_{s(d)})
\]

We reparameterize to work with the unconstrained

\[
\omega_t = \left( \log \left[ \frac{\pi_{t1}}{\pi_{tK}} \right], \ldots, \log \left[ \frac{\pi_{t(K-1)}}{\pi_{tK}} \right] \right)'
\]

In order to capture dynamics in \( \pi_t \) and to borrow strength from neighboring time periods, we assume that \( \omega_t \) follows a Dynamic Linear Model (DLM) (West and Harrison, 1997; Cargnoni et al., 1997; Martin and Quinn, 2002). Specifically,

\[
\omega_t = F_t' \eta_t + \epsilon_t \quad \epsilon_t \sim N(0, V_t) \quad t = 1, \ldots, T \tag{1}
\]

\[
\eta_t = G_t \eta_{t-1} + \delta_t \quad \delta_t \sim N(0, W_t) \quad t = 1, \ldots, T \tag{2}
\]

Here Equation 1 acts as the observation equation and Equation 2 acts as the evolution equation. We finish this prior off by assuming prior distributions for \( V_t, W_t, \) and \( \eta_0 \). Specifically, we assume
$W_t = W$ for all $t$ and $V_t = V$ for all $t$ in which Congress was in session with $V$ and $W$ both diagonal and

$$V_{ii} \sim \text{InvGamma}(a_0/2, b_0/2) \quad \forall i$$

and

$$W_{ii} \sim \text{InvGamma}(c_0/2, d_0/2) \quad \forall i$$

We assume

$$\eta_0 \sim \mathcal{N}(m_0, C_0)$$

For days in which Congress was not in session we assume that $V_t = 10I$. We have found that this helps prevent oversmoothing.

Note that what we have done is to specify a hierarchical prior that factorizes as:

$$p(z, \omega, \eta, V_t, W_t) = p(z|\omega)p(\omega|\eta, V_t)p(\eta|W_t, m_0, C_0)p(V_t)p(W_t)$$

In what follows we specify $F_t$ and $G_t$ as:

$$F_t = \begin{pmatrix} I_{K-1} \\ 0_{K-1} \end{pmatrix} \quad t = 1, \ldots, T$$

$$G_t = \begin{pmatrix} I_{K-1} & I_{K-1} \\ 0_{K-1} & I_{K-1} \end{pmatrix} \quad t = 1, \ldots, T.$$
method is the fact that it can be used without any manual coding of documents it can also be
used in semi-supervised fashion by constraining some elements of $Z$ to be 0 and 1 in order to force
particular documents into particular topic categories.

Appendix A presents details of the Expectation Conditional Maximization (ECM) algorithm
used to fit this model.

3 Data and Measurement

The data used here are drawn from the United States Congressional Speech Corpus (Monroe
et al., 2006) developed under the Dynamics of Political Rhetoric and Political Representation
Project (Monroe et al., 2005). The input data used here consist of frequency counts of word stems
in Senate speech from the 105th-108th Congresses (1997-2004). These are generated through several
steps. The raw source for the data is the electronic version of the (public domain) United States
Congressional Record served by the Library of Congress on its THOMAS system (The Library of
Congress, N.D.) and generated by the Government Printing Office (The United States Government
Printing Office, N.D.). The Congressional corpus covers both chambers for the time period from the
101st Congress (1988-) to the present; this paper uses the subset of these data from the 105th-108th
Senates only.

The basic units of the electronic version of the Congressional Record are “html documents”. The Record
is subsetted into three major components, one for the Senate and two for the House
of Representatives (one of which covers “extensions”, or speeches inserted into the record without
being given on the floor). These are subsetted by day. Days are subsetted into the html documents,
corresponding roughly to titled subsections in the printed Record. This subsetting process is not
regularized and html documents can contain zero, one, or several speakers, discussing one or more
items or topics. There are roughly 9,000 of these documents per year in the Senate; there are 71,181
documents from 1997-2004 used here.

We then parse these html tokens to create a tagged XML version. Each paragraph is tagged as
speech or nonspeech, with speeches further tagged by a speaker id. An excerpt is shown in Figure 1. Unlike some parliamentary records (such as the Australian *Hansard*), the *Congressional Record* is not already tagged in this way, so it is a fairly elaborate and time-consuming process to identify patterns that indicate the start and stop of speech and that identify the speaker. Changes in formatting over time, misspellings, typographic errors, and similar inconsistencies are surprisingly common.

For the present paper, we define our unit of analysis as a *speech document*. A single observation of a speech document consists of all text tagged as speech with the same speaker within a single html document. This operationalization will adjoin in one speech any parts separated by the intervention of another speaker (as with the question in Figure 1) and will not recognize as separate any distinct speeches by the same person within the same html document. It will also separate any speech that continues across two html documents, although we are unaware of any examples of this beyond procedural speakers and leaders in very long debates.\(^\text{11}\) There are roughly 15,000 speeches observed within a given year of the Senate; our data here are based on 118,065 such speech documents from 1997-2004.

Each speech is then stripped of capitalization and punctuation following a tuned set of rules (details in Monroe et al. (2006)) that allows some non-alphanumeric information to remain (e.g., *don’t*, *9/11*, *$17*). The resulting strings of words are then *stemmed* using the Porter’s Snowball II stemmer (for English), as is standard in natural language processing applications (Porter, 1980, N.D.).\(^\text{12}\)

\(^{11}\)The very small number of such cases result in one long “speech” being divided into two shorter ones.

\(^{12}\)This attempts to reduce variant word forms – plurals, present participles, etc. – into common stems. The Porter stemmer is aggressive, erring on the side of “over-stemming.” It allows for user-tuned exceptions, but we have made no attempt to identify any. As a result, the reader may note a few minor oddities in the tables below. For example, *IRS* has been lower-cased to *irs* and then “stemmed” to the “singular” *ir*. This makes absolutely no difference, however, as “ir” means only “IRS.” For our purposes, the instances where this actually merges two words that should not be merged, e.g., *his* and *hi*, are vastly outweighed by the increased leverage from merging words by stem. We can, for example, leverage the shared topical inference from words like *abort*, *aborts*, *aborted*, *aborting*, *abortion*, *abortions*, *abortionist*, *abortionists*, rather than treating them as independent words (with different parameters to estimate). Further intuition for this can be easily confirm with Google searches, where stemming is used. A plural keyword will return pages matching the singular and searches for the present participle or gerund return matches without the “-ing”. This makes it much more likely that pages of relevant and similar content will be identified, similar to our goals here.
We then count the stems in each speech. These stem counts per speech are the inputs to the inferential model. The vast majority of unique stems are infrequently used. Such stems provide little information to the model while increasing the computational demands. For our purposes here, we filter out any stems that occur in fewer than 0.05% of speeches. The entire 16-year Senate corpus contains roughly 240,000 unique stems; our eight-year period from 1997-2004 contains 154,351 unique stems. After filtering, our estimations here are based on 3,807 unique stems with a total of 73,990,638 stem observations.

4 Results from 105th-108th U.S. Senate

We fit numerous specifications of the model outlined in Section 2 to the 105th-108th Senate data. In particular, we allowed the number of topics $K$ to vary from 3 to 60. For each specification of $K$ we fit several models using different starting values. Mixture models, such as that used here, typically exhibit a likelihood surface that is multimodal. Since the ECM algorithm used to fit the model is only guaranteed to converge to a local mode, it is typically a good idea to use several starting values in order to increase one’s chances of finding the global optimum.

As $K$ moved up to about 20 the rough substantive outlines of the clusters began to take shape with clusters that corresponded to topics such as judicial nominations, education, use of military force, health care, etc. However, until we moved $K$ up to about 40 there appeared to be some lumping together of categories that made interpretation difficult. The results we present below are based on the 42 topic fit with the highest maximized log posterior.

4.1 Interpreting the Topics

Table 1 provides our substantive labels for each of the 42 clusters, as well as descriptive statistics on relative frequency in the entire dataset. We decided on these labels after examining $\beta_k$ and also reading a modest number of randomly chosen documents that were assigned a high probability of being on topic $k$ for $k = 1, \ldots, K$. We discuss each of these procedures in turn.

In order to get a sense of what words tended to distinguish documents on a given topic $k$ from
documents on other topics we examined both the magnitude of $\beta_{kw}$ for each word $w$ as well as the weighted distance of $\beta_{kw}$ from the center of $\beta_{-kw}$. The former provides a measure of how often word $w$ was used in topic $k$ documents relative to other words in topic $k$ documents. A large positive value of $\beta_{kw}$ means that word $w$ appeared quite often in topic $k$ documents. The weighted distance of $\beta_{kw}$ from the center of the $\beta_{-kw}$, which we operationalize as:

$$r_{kw} = \frac{\beta_{kw} - \text{median}_{j \neq k}(\beta_{jw})}{\text{MAD}_{j \neq k}(\beta_{jw})},$$

provides a measure of how distinctive the usage of word $w$ is on topic $k$ documents compared to other documents. To take an example, the word “the” always has a very high $\beta$ value as it is very frequently used. However, it is used roughly similarly across all of the topics so its value of $r$ is quite close to 0. We combine these measures by ranking the elements of $\beta_k$ and $r_k$ and adding the ranks. This combined index gives us one measure of how distinctive word $w$ is for identifying documents on topic $k$. Table 2 provides the top keys for each topic as determined by this index.

Inspection of these tables produced rough descriptive labels for all of the clusters. After arriving at these rough labels we went on to read a number of randomly chosen speech documents that were assigned to each cluster. In general we found that, with the exception of the procedural categories, the information in the keywords (Table 2, extended) did an excellent job describing the documents assigned to each (substantive) topic. However, by reading the documents we were able to discover some nuances that may not have been apparent in the tables of $\beta$ values, and those are reflected in the topic labels and clarifying notes of Table 1.

In general, the clusters appear to be highly homogeneous and well-defined. Figures 2 and 3 present excerpts from speech documents that had high (over 0.9999) probabilities of being on what we label as ‘Law and Crime 1 [Violence]’ and ‘Education’ topics. Reading these excerpts we see that these speech documents do appear to be on the topics of crime and education respectively. Senator Murray’s speech is particularly interesting because it uses words such as “terrorism”, and

\[\text{Longer lists of keywords and index values are provided in the web appendix ?.}\]
“medical” and “psychological” that might seem out of place in an education speech. Yet, because our model takes into account the overall pattern of word usage in a document we still correctly classify this document as an education speech. It is worth noting that this technique will extract information about the centroid of the cluster’s meaning and lexical use. There will be speeches that do not fall comfortably into any category, but which are rare enough not to demand their own cluster.

Reading some of the raw documents also revealed some additional meaning behind the clusters. For instance, two of the clusters with superficially uninformative keywords turn out to be composed exclusively of pro forma “hobby horse” statements by Senator Jesse Helms about the current level of national debt, and by Senator Gordon Smith about the need for hate crime legislation.

The $\beta$ parameters identify words that reliably separate one topic area from all others, for the time period under study and for the Senate as a whole. For different time periods, particular terms would not exist (mccain-feingold). While some would be less evocative of particular subjects (e.g., iraq), others would be more so. More important, our approach does not demand that all legislators talk about all topics in the same way. To the contrary, there is typically both a common set of terms that identifies a topic at hand (as shown in Table 2) and a set of terms that identify particular political (perhaps partisan) positions, points of view, frames, and so on, within that topic. For example, consider the keywords for the Judicial Nominations topic, shown in Table 3. The table shows an expanded list of the topic-identifying keywords as given in Table 2 along with the keywords that most indicate, within speeches during the 108th Senate (2003-4), that the speaker is either a Democrat or a Republican. These are derived through the “rhetorical ideal point estimation” procedure described in detail in Monroe and Maeda (2004). We note here only the extremely high face validity of the partisan keys and the distinction between the types of keys. There can be words used by all in common that suffice to classify documents by topic, while there are simultaneously words within the topic that identify the recent partisan conflict in the Senate. Democrats argue that

\footnote{And which would lead to double-counting in an information-retrieval approach.}
an abusive Republican leadership insists on its agenda to pack courts with ideological judges like Scalia, while Republicans counter that their Democratic opponents, emboldened by Bork, use the supermajority filibuster tactic to avoid up-or-down votes on judicial nominees, in order to block qualified judges (cum laude graduates of Harvard and Columbia) who fail their pro-tax, pro-abortion litmus tests.

4.2 The Substantive Structure of the Rhetorical Agenda

The $\beta_k$ vector not only helps identify the internal meaning of each topic (at its centroid), but provides a means of identifying how the topics relate to one another. The face validity of these relationships provides further support for the claim that we are identifying topics with coherent substantive meaning. Moreover, this informs efforts to identify subtopics of larger topics.\textsuperscript{15} Perhaps more important, this can inform literatures about the theoretical relationships among policy areas and the proper empirical approaches to policy and issue coding.

In order to get some sense of how the 42 topics related to each other we performed agglomerative clustering of $\beta_1, \ldots, \beta_{42}$. Agglomerative clustering begins by assigning each of the 42 vectors to its own unique cluster. The two vectors that are closest to each other (by Euclidean distance) are then merged to form a new cluster. This process is repeated until all vectors are merged into a single cluster. The results of this process are displayed in the dendrogram of Figure 4.\textsuperscript{16} Roughly speaking, the lower the height at which any two topics, or groupings of topics, are connected, the more similar are their word use patterns in Senate debate.

Reading Figure 4 from the bottom up provides information about which clusters were merged first (those merged at the lowest height). We see that topics that share a penultimate node share a substantive or stylistic link. Some of these are obvious topical connections, such as between the two health economics clusters or between energy and environmental regulation. Some are more subtle. For example, the “Environment 1 [Public Lands]” category, which is dominated by issues related

\textsuperscript{15}This is especially useful given the relatively arbitrary modelling choice of $K$.

\textsuperscript{16}The order of topics given in Tables 1 and 2 is as determined here; the labels were determined prior to the agglomerative clustering.
to management and conservation of public lands and water, and the “Commercial Infrastructure” category are related through the common reference to distributive public works spending. Both contain the words *project* and *area* in their top 25 keys, for example. The “Banking / Finance” category and the “Labor 1 [Workers]” category discuss different aspects of economic regulation and intervention, the former with corporations and consumers, the latter with labor markets. Other connections are stylistic, rather than necessarily substantive. The symbolic categories, for example, all have *great, proud,* and *his* as keywords.

We can also read Figure 4 from the top down to get a sense of whether there are recognizable rhetorical metACLusters of topics. Reading from the top down we see clear clusters separating the housekeeping procedural, hobby horse and symbolic speech from the substantive policy areas. The more substantive branch then divides a cluster of conceptual and Constitutional issues from the more concrete policy areas that require Congress to appropriate funds, enact regulations, and so on. Within the concrete policy areas, we see further clear breakdowns into domestic and international policy. Domestic policy is further divided into clusters we can identify with social policy, public goods and infrastructure, economics, and “regional”.17 Note that what binds metACLusters is language. The language of the Constitutional grouping is abstract, ideological, and partisan. The social policy grouping is tied together by reference to societal problems, suffering, and need. The public goods / infrastructure grouping is tied together both by the language of projects and budgets, as well as that of state vs. state particularism. The most interesting metacluster is the substantively odd “regional” grouping of energy, environment, agriculture, and trade. Exploration of the language used here shows that these are topics that divide rural and or Western senators from the rest – distributive politics at a different level of aggregation.

Note also how some substantive issues are *separated* in perhaps unexpected ways. Health policy, for example is split into four categories. Speeches that emphasize the economic aspects

---

17 The “health economics” grouping is an outlier because of the artificial unidimensionality of the dendrogram. The clusters mixes the language of the social “health” category with that of the economic “debt / deficit” category, but these are not next to one another in the best one-dimensional ordering of the topics.
of health care (costs of Medicare, prescription drugs, insurance, provision, etc.) are placed near other economic categories. Speeches that emphasize the medical aspects of health care (prevention, research, disease, suffering) are placed within the other social policy categories. Abortion is placed with the (highly partisan) constitutional grouping. Similarly, military topics can be found in the form of discussion of war under the defense/international grouping, but also in the distributive grouping (e.g., base closures, homeland defense, VA hospitals) and in the symbolic grouping (e.g., praising soldiers killed in service, funding the WWII Memorial).

This approach has the potential to inform ongoing debates about how to characterize the underlying political structure of public policy. Whether such characterization efforts are of interest in and of themselves—we would argue they are—is not of as much relevance as the fact that they are necessary for understanding dimensions of political conflict (Clausen, 1973; Poole and Rosenthal, 1997), the dynamics of the political agenda (Baumgartner and Jones, 2002; Lee, 2006), the nature of political representation (Jones and Baumgartner, 2005), or policy outcomes (Lowi, 1964; Heitschusen and Young, 2006; Katzenelson and Lapinski, 2006). Katzenelson and Lapinski (2006) provide an eloquent defense of the exercise and a review of alternative approaches, along with their own.

There have been three basic approaches. One is to build up a classification system from first principles rooted in the substantive content (be it actual or theoretical) of policy. This is the approach of Lowi (1964), Katzenelson and Lapinski (2006), and others. Two policy areas are related, for example, if they both address the boundaries of state-citizen relationships, management of resources, and so on. A second is inductive, analyzing legislator behavior for patterns. This is the approach, using roll-call votes, taken by Clausen (1973), Poole and Rosenthal (1997) and many others. Policy areas are related if they generate similar patterns of voting behavior. The third approach, that of Baumgartner and Jones (2002) is somewhat of a hybrid, starting with substantive categories roughly matching Congressional committee and executive branch departmental structure (e.g., health, education, defense, etc.) and then inductively adapting to achieve a mutually exclusive
and exhaustive set of categories that can be coded with high reliability. Their approach has been
the most widely adopted, picked up in studies of legislative agendas (Adler and Wilkerson, 2006)
and comparative policy-making (Baumgartner et al., 2006).

Our approach is closest in principle to the roll-call studies, being purely inductive. For the
purposes of identifying relationships among policy areas, we have three distinct advantages. The
first is that language gives us two features of similarity to examine, rather than one. Roll-calls are
tied together – deemed similar or dissimilar – only by the legislators that vote for and against them.
Words can be tied together both by legislators that use them and by their use in conjunction with
other words. The second advantage is that words convey meaning directly, evoking immediate clues
to the commonality of collections of words (as seen with the keywords). The third advantage is the
ability to escape the spatial “dimensionality” constraints of roll-call analysis. A good example of
this is distributive politics. Truly distributive politics in the Senate would be each-state-for-itself,
and would require a 49-dimensional simplex to represent spatially. It is essentially impossible for
roll-call voting records to be rich enough to capture that. Our approach allows us to recognize that
the grouping is dominated by terms like project, area, outlay, FY, earmark, state names, and so on
that flag the distributive nature of the debate.

Figure 5 shows a partial crosswalk between the 29 substantive topics of the 42 topic model, the
implied rhetorical meta-clustering, the meta-clustering into six major groupings, and the upper-
level classifications of Katzenelson and Lapinski (2006) (Tiers 3 and 2), Jones and Baumgartner
(2005) (Super- and major categories), Lee (2006), and Clausen (1973). A filled cell in the diagram
indicates some shared content. Multiple filled cells in a row indicate the grouping on the left
captures more than one rhetorical topic cluster; multiple filled cells in a column indicate that a
rhetorical topic spans more than one of the groupings in a classification.

There are clearly differences. For example, ours is the only of these to classify foreign trade as
a domestic, rather than international, topic. This reflects the reality that trade is debated in terms
of its domestic economic impacts, not its diplomatic or strategic implications. This suggests the
politics of trade are more similar to the politics of agriculture than the politics of war or human rights.

Another example can be found in the public goods grouping, where we find the CIA (an international category for most), the FBI (put with crime by most), and immigration (international for Clausen, “Sovereignty / membership and nation” for Katzenelson and Lipinksi, labor and employment for Baumgartner and Jones) among others. This reflects post-9/11 debate and policy shift to reorganize multiple agencies into a new Department of Homeland Security. Quickly, the politics of DHS came to focus on particularistic concerns about the geographic distribution of its funding, more closely resembling the terms of competition surrounding issues like transportation appropriations (in our commercial infrastructure topic) or military base closures (in our armed forces infrastructure topic).18

Of course, much of this is time-bound. A time window prior to 9/11 would not have found so many issues tied together by post-9/11 politics. A time window extending into 2006 would surely identify immigration a category (or more) of its own. It might then be placed not with the INS and Border Patrol alongside other DHS debates, but plausibly within the economic meta-cluster for its labor market implications, the conceptual meta-cluster for its citizenship implications and English-only frames, or both. And, of course, we see nothing resembling historical categories like “frontier settlement” that appear in the Katznelson and Lapinski (2006) scheme.

This is emphatically not an argument that other classification systems are “wrong”. Each is designed to serve a different particular theoretical or empirical purpose and each does so, and several have been designed to be repurposed as desired. The PAP approach, for example, identifies more than 200 fine-grained ‘minor topic” areas. While these are aggregated in the primary Baumgartner and Jones work to 19 or 20 major categories and, in one case, six supercategories (Jones and

18Satirical talk-show host Jon Stewart summarized the debate: “Tuesday, the Senate voted to continue distributing a significant portion of Homeland Security dollars equally among the states, rather than by likelihood of attack. Bad news for states like New York and California; good news for small states like Wyoming, which only has one high-risk target: the popular tourist attraction ‘The World’s Largest Pile of Homeland Security Money.’ ” The Daily Show, July 14, 2005.
Baumgartner, 2005, p. 240), the presence of the minor categories allows any arbitrary re-aggregation the analyst might deem appropriate. It would be straightforward to re-aggregate their data to match our groupings or any other, if one were so inclined.

We make no strong claims for the particular inter-relationships among categories that we find here. There is an artificial unidimensionality to Figure 4 that obscures the secondary relationships among categories – across military or health policy topics, for example. We fully expect this basic picture to change with further analysis using much larger $K$, data for different periods of times, and data for the House of Representatives. We do think, however, that we have strong evidence that politicians talk about these topics differently than policy analysts and political scientists do, which likely has important implications for how we study policy-making.

4.3 Policy and Issue Dynamics

Our efforts here are aimed most directly at understanding the dynamics of issue debate and policy agendas. Our primary frame of reference here is PAP (Baumgartner and Jones, 1993, 2002; Jones and Baumgartner, 2005). Our approach offers direct leverage on many of the questions they address, complementing the understanding of long-term dynamics made possible by their studies with an understanding of short-term dynamics made possible by ours. When exactly – to the day – have particular topics been on the agenda, how has attention been distributed among topics, what dynamic processes govern agenda content, and what dynamic links, if any, exist between the political agendas of elites and the public? Our approach allows us to explore these questions in ways that have not previously been possible.

4.3.1 Descriptive Summaries of Agenda Dynamics

First and foremost, our model estimates a number of quantities of interest directly that provide useful descriptive summaries of the the distribution of attention to topics in the otherwise overwhelmingly large Congressional Record. Baumgartner and Jones (1993) give a compelling defense for why such “simple graphical illustrations of the data” illuminate important aspects of agenda
dynamics even where data limitations (time series with less than 50 annual observations) limit more elaborate statistical analysis (Appendix B), and this continues to be an important engine in the more recent PAP research (Baumgartner and Jones, 2002; Jones and Baumgartner, 2005). In addition to this primary descriptive purpose, such illustrations serve here to reinforce the face validity of our approach as well as to make crystal clear that there are dramatic subannual dynamics in the political agenda. We illustrate this with a few interesting examples from our “Constitutional / Conceptual” metachipuster.

Consider the ‘Judicial Nominations’ category. We show in Figure 6 the estimated probability that, on any given day, a document will be in this category. The apparent trend in this probability over time presents some evidence that the nomination process has been given increasing attention from 1997 to 2004. In general, the fraction of speeches on ‘Judicial Nominations’ appears to increase just before the 2002 election when it looked likely that the Republicans would retake the Senate and thus allow President Bush more freedom to appoint conservative jurists. It continues to grow and experience outbursts of speech (and partisan rancor) as Senate Democrats exercise their one remaining institutional piece of power, the filibuster. In fact, the record-holding debate in our data – the most words on one issue in a single “day” of the Congressional Record – was on this topic and actually spread over two days in the real world. This was the November 12, 2003 “Justice for Judges” marathon, in which Senator Rick Santorum organized a 39-hour debate on four federal judge nominations, attempting unsuccessfully to overcome a Democratic filibuster and get “up-or-down votes.” Our data show 87 speeches and over 230,000 words spoken in the Judicial Nominations category on November 12.

The other major judicial topic, ‘Supreme Court / Constitutional’, exhibits very different dynamics. These can be seen in Figure 7. This series contains the second largest topical outburst in our data: the conclusion of the Clinton impeachment trial (February 12, 1999, 78 speeches, 171,000 words). This series also includes a number of other constitutional debates related to the Supreme Court, the Department of Justice and Attorney General, and possible constitutional amendments.
For example, the next largest spikes appear in January 2001 (Ashcroft nomination) and July 2004 (Federal Marriage Act).

Figure 8 displays the marginal probability of seeing a document in the ‘Abortion’ category over time. Here we see that attention to abortion in the Senate waned from the late 1990s to 2004. In the 105th Senate, there were instances where approximately 6% of the documents were on the ‘abortion’ topic. Aside from a single moderate spike during the partial birth abortion debate in March 2003, the amount of abortion rhetoric during the Bush presidency has been generally quite low and stable around 1%. We suspect that Senate Republicans introduced the abortion issue more frequently during the Clinton presidency as a largely symbolic device. It certainly suggests a bit of “stickiness” to the abortion issue – a tendency to stay on or off the agenda for long stretches of time, an issue we explore in the next section.

4.3.2 Evaluating Subannual Agenda Dynamics

The Baumgartner-Jones work demonstrates convincingly that dynamics are a (and perhaps the) central theoretical concern for scholars of the policy process. The key to understanding the policy process is understanding the variety of mechanisms by which policy agendas come to experience trends, stochastic shocks, positive and negative feedback. This motivates the massive data collection effort of PAP, with which they and many others have examined dynamics occurring at an annual level over the postwar era. While informing questions about long-term dynamics, annual data necessarily mask any political dynamics at smaller time scales. These can be important, including including election runups, budgetary cycles, seasonality, response to events, temporal causality among institutions and different policy arenas (e.g., the president, Congress, media, public opinion), and so on. We take the fist step here of demonstrating that we can, with our approach, illuminate subannual dynamics in the rhetorical agenda that (a) are of interest, and (b) are beyond the reach of other approaches.

A first concern is to examine and compare the dynamics of the relative distribution of attention
to each topic area. Roughly speaking, we evaluate our posterior estimate of the (log-odds) of words in documents we have classified in a particular category as a univariate time series using conventional Box-Jenkins methods. For the daily data, we find that an AR-7 (autoregressive lags up to 7 days) with monthly dummy variables, captures most of the relevant daily dynamics. Once estimated, each series can be interpreted for its persistence and volatility.19

Consider first the issue of persistence. In our data, this reflects the “stickiness” of surprises, or memory, in particular issue areas. That is, once something causes a topic to be discussed more (or less) than expected on a particular day – to be disturbed from its equilibrium – how long does it take to return to equilibrium? This type of notion is fundamental to Baumgartner and Jones’ notions of punctuated equilibrium in the policy agenda, as well as the theoretical alternatives (e.g., incrementalism, issue cycles, garbage can processes, etc.).

Figure 9 shows this phenomenon, as derived from the AR-7 model on words, for a selection of topics. Generally, shocks in the rhetorical agenda dissipate away from the system within two to three weeks.20

Some issues have essentially no memory whatsoever. The most extreme example in our data

---

19Several refinements of this process could be explored in the future. First, the data used in this section are the output from our topic categorization model. One useful extension would be to incorporate the uncertainty surrounding topic categorization into the dynamic analyses. Second, we model the log-odds of seeing a topic as following a specific dynamic specification that is invariant across topics. In many ways, this makes it more difficult to discern unique topic dynamics from the data. Future recursive estimation of the topic categorization model and the topic dynamics, allowing the persistence and memory of the topic to diverge could be attempted in the future. A second, more practical step, would be to alter our prior specification for the local linear trend model and analyze the effect this has on the dynamic results. Third, the data are inherently compositional (Aitchison, 1986). Failure to incorporate all regressors in all models - most notably the lags of all categories on all categories - biases our results. This is probably slight for 42 categories, but a more principled model would be something like the compositional VAR described (CVAR) by Brandt et al. (1999). However, a 42-equation CVAR model is impractical at this point. Fourth, even with the logratio transformation, the errors are not Gaussian. There is clearly zero-inflation, an increased probability that a topic will not be addressed at all on any given day. Finally, ARIMA models are known to be sensitive to even small specification changes. ARIMA(6,0,0) and ARIMA(8,0,0) models did not yield qualitatively different results. As more data becomes available annual and election-timing multiplicative seasonality should be examined in depth. None of these should have dramatic impact on the broad qualitative observations we make here. More to the point, our main purpose here is to demonstrate what is possible. Scholars who would like to try a different approach to specification, now have the data and tools to do so.

20It belabors to say that this would be impossible with annual data. Note, however, that our data series are not long enough to capture the long-term dynamics, on the order of presidential regimes, seen in the Baumgartner-Jones work. This is an important reason we view the approaches as completely complementary. Note also that, given sufficient computing power, only data availability prevents such long-term analysis with our approach. Should the Congressional Record be digitized for a longer period, we will explore this. Several of our comparative datasets under development in the larger project do in fact span 3-18 decades.
is the “Symbolic [Sports]” category. These speeches are made to congratulate home-state sports champions. No one stands up the next day to rebut or get equal time. As shown in Figure 9, shocks in this series are completely ephemeral, disappearing completely by the next day. There are more substantive ephemera as well. The topic dealing with violent crime is a good example. Shocks in attention here are largely speeches about events, comments on the news. They do not lead to sustained discussion, however, losing 95% of attention in a day and disappearing a week later.

Other agenda items have much more persistence. The extreme examples in our data are the two one-person hobby horse categories: Jesse Helms’ almost daily speech on the debt and deficit until his retirement and Gordon Smith’s almost daily speech on hate crime starting in 2001. Here, “persistence” means exactly that. Once they started, they kept going (until, in Helms’s case, he retired and stopped for good). The next most persistent topic is abortion. This likely reflects at least three factors: polarization (speeches on one side draw response on the other), intrapartisan unity (party leaders can move blocs of speakers to talk or not talk about the issue at once), and electoral cycles (it receives sustained attention in election runups). Many categories are persistent because circumstances demand the Senate’s sustained attention in some institutional role (e.g., judicial nominations, impeachment, war), or because there is a sustained debate about a major policy proposal (e.g., tax reform, intelligence reform, bankruptcy reform, education reform, prescription drug coverage, campaign finance reform).

The second major dimension to dynamics is the structural volatility of each topic. Simply put, how variable is the topic – how predictable is discussion of the topic from day to day? We can summarize this as typical error in the one-day ahead forecast from our AR-7 model. Figure 10 shows the persistence and volatility of all 42 topic areas. In the upper left, we see ephemeral commentary on unpredictable topics (sports, crime, health care topics of the day). In the lower right, we see topics that appear and disappear predictably, and stay on the agenda when they do (highly partisan and electoral issues). In the upper right, we see the issues that can surprise, and
stick on the agenda when they do. This includes major policy proposals (No Child Left Behind in education), major economic shocks (employment), and major world events (9/11 followed by intelligence reform and war). In the lower left, we see the issues that have predictable and brief bouts of attention. Here we find technocratic policy areas that require Congressional attention, either at a steady rate or on an annual cycle, but have low (partisan) political salience during the period, with examples including wildlife conservation, internet regulation, and human rights.

4.3.3 Intervention Analysis: 9/11

In addition to the internal dynamics of attention, we can examine the dynamic impact of exogenous factors, like events. The most dramatic event in our data is 9/11. This is an event that occurred on a single day, and which we know had both immediate and long-term effects on both rhetoric and policy in American politics. We can incorporate such “interventions” easily into models like the one used in the previous section and get direct measures of the rhetorical impacts. This is a sharp test of face validity for the model – if 9/11 effects, of all things, are indiscernible or implausible, we have little reason to expect we will find anything more subtle. Again, such effects are impossible to detect without data at a sufficiently fine temporal grain.

The left two columns of Table 4 summarize the results of a standard intervention analysis. The Senate did not convene on 9/11.\textsuperscript{21} When it convened on September 12, the “surprises” in rhetoric – those that were difficult to predict on September 10 – are placed in the “international affairs / arms control” and “symbolic – military” categories. This partially reflects the uniqueness of the speeches in the immediate aftermath. Not being plentiful enough across the entire dataset, they do not form a category of their own but are placed with the closest topical mode. Over the full eight years, the symbolic military category is dominated by speeches honoring fallen soldiers. On September 12, the speeches honoring policemen, firemen, and other rescue workers are in this category. The arms control category is dominated, over the full eight years, by discussion of weapons

\textsuperscript{21}The data clearly show an abrupt shift after 9/11, so a second-order transfer function was unnecessary. One pulse and one permanent intervention variable allows us to investigate both the initial and long-term reactions to terrorist attacks. The return to a (potentially) new equilibrium was adequately modelled by the autoregressive dynamics.
of mass destruction, nuclear nonproliferation, international military organizations (e.g., NATO), and similar topics. Speeches from September 12 that declared the 9/11 attacks an act of war (but from an as yet unknown assailant) – use of force / attacks by someone else – are placed by the model in this category. Also increasing that day, but not reaching statistical significance, were speeches categorized in the other international affairs category, in defense (use of force), in armed forces infrastructure, and in intelligence, as well as in other symbolic categories. There is also more procedural speech reflecting the difficulty in arranging sufficient time for everyone who wanted to speak.

The long-term effects (Table 4) include increases in the expected war / military / terrorism categories: use of force, intelligence, and the three military categories (manpower, infrastructure, and symbolic). There are also spurious effects, typical in such models. For example, Gordon Smith started his hate crime speeches in April of 2001 and continued through the end of the period. So, while the equilibrium level of such speeches for 9/12/2001-2004 is indeed higher than the equilibrium level for 1997-9/11/2001, it doesn’t have anything do with the attacks. Keeping in mind this caveat, the long-term post-9/11 equilibrium seems to involve significantly less attention to a few social issues (e.g., child protection, education) as well as a few of the Republican wedge issues used in opposition to Clinton (e.g., abortion, constitutional).

To the extent something might reflect a Clinton-Bush change, rather than a post-9/11 change, we can separate anything that was discernible in the nine months starting the Bush term. This is difficult, however, since this period also includes the party switch of Senator Jeffords. This altered majority control of the Senate and also presumably altered the pattern of rhetoric. Nonetheless, if we include a Bush regime variable as well, we can separate out some pre-9/11 agenda effects due to the presidency change from those that followed 9/11.

The second half of Table 4 summarizes these results. Here we see two substantive categories considered to have significantly increased attention in the first year of the Bush presidency, but

\[ \text{Equation} \]

\[ \text{Comment} \]

22Specifically, all apparent effects could be related to other structural changes with a similar time pattern, especially the change of presidential regime from Clinton to Bush and economic softening.
prior to 9/11: energy and military personnel. Moreover, there is significant decreased attention to the two wedge issues (abortion and constitution), to use of force, as well as to child protection and agriculture. This brings the true post-9/11 effects – especially increased attention to use of force, intelligence, symbolic military speech – into sharper relief.

Some of these estimated effects are large enough to be discerned through the descriptive plots. Figure 11 shows the probabilities of randomly chosen documents being on one of three foreign policy categories: “Use of Force”, “Arms Control”, and “Diplomacy / Human Rights”. Prior to October 2002 (when Congress authorized the use of force in Iraq) Senate speech was roughly evenly split between these three topics. After this point, the overwhelming majority of Senate speech on these three topics falls into the ‘Use of Force’ category. In fact, the debate authorizing the use of force in Iraq is the third largest topical outburst in our data (October 10, 2002, 55 speeches, 122,000 words). While it shouldn’t be surprising that this increase occurred, it is a bit more interesting to note that the attention paid to the other two categories diminishes somewhat during this same time.

It is also interesting to see how $\pi_{Symbolic-Military}$ tracks with $\pi_{Defense-UseOfForce}$. We plot both series from August 2001 to May 2003 in Figure 12. This plot seems generally consistent with what is known of the lead-up to the Iraq war. As noted above, we see a large increase in symbolic speech immediately after 9/11. At this point there is very little talk of foreign intervention. Talk of intervention begins to increase a bit in spring 2002 and then there is large upswing in October 2002 when Congress authorized President Bush to use force against Iraq. Things are relatively quiet until February when, as invasion appears imminent, the amount of Senate speech on foreign intervention and Iraq increases. During combat operations in March 2003, talk of military intervention falls off and related speech becomes dominated by symbolic tributes to soldiers and the military.
4.3.4 Macro-Representation in Rhetoric

Students of democratic representation are interested, in part, in the extent to which elected officials follow or lead their citizens. If representatives respond to public opinion, it can be seen positively as responsiveness (Stimson et al., 1995) or negatively as pandering; if public opinion moves in response to representatives, it can be seen positively as leadership (Jacobs and Shapiro, 2000) or negatively as the result of manipulative framing.

Annual data are unlikely to pick up temporal change at time-scales appropriate to this question (or, more accurately, cannot be used to assess what time scale is appropriate for this question) and are therefore limited in application largely to measures of simultaneous “congruence” in policy attention (Jones and Baumgartner, 2005). Stimson et al. (1995), on the other hand, have finer time scales, but only highly aggregated measures of legislative and citizen preference that cannot be disaggregated into, say, attention across policy topics. Our data are fine enough, across time and policy, to evaluate such things and provide important complements to these prior approaches.

As a first exploration, we examine leadership/followership of legislative rhetoric with issue salience, as measured by the Gallup “most important problem” question. The Policy Agendas Project has collected these data and coded answers according to their 19+ topic coding schema.23 These are available almost monthly for our time period and are smooth enough to allow for reasonable interpolation across missing months. In turn we aggregate our measures up to the monthly level to allow for direct comparison.

To gain insight into the question, we examine the Granger causality between pairs of MIP-poll series and attention in relevant rhetorical categories (Granger, 1969; Sims, 1972). This technique examines two univariate time series to determine the extent to which the first can be used to reliably forecast the second, the second to forecast the first, both, or neither. If one series can forecast a second, and not vice versa, it means that movements in the first reliably precede those of

---

23We are grateful to Frank Baumgartner for providing these data and Amber Boydstun for help in adapting them to our purposes.
the second. This is probably a necessary condition for the first to be causally prior to the second. But it is certainly not sufficient. Both may be caused by a third phenomenon, with the first simply responding more quickly than the second. Furthermore, the first may simply anticipate the second, preceding it without causing it.

Consider first foreign policy. Figure 14 illustrates the series used with a few Defense/International categories of MIP and rhetoric. Table 5 provides the results of the tests. We find that, over this time period, legislative speech in all categories related to the military and terrorism does in fact lead / anticipate the public salience of foreign policy. There is minimal evidence of causality in the nonmilitary aspects of foreign policy, including trade (consistent with our earlier finding that rhetoric about trade is based more on domestic economic implications than on foreign relations) and minimal evidence of the public leading the Senate in any of these categories.

In short, it seems very likely that the public calls for attention to defense / international issues only after elites are already paying attention to them. This could be because elite attention raises public attention in this area. It could also be that presidential priorities or the external realities of war, bombing, and the like receive swifter attention from the Senate than the public. It could even be that the Senate pays attention quickly to such issues because they know the public will soon demand they do so. In any case, it seems very unlikely that the Senate pays attention to defense and international affairs only after it becomes salient in public opinion. We can, in other words, provide strong evidence against pandering/responsiveness in international affairs.

At a lower level of attention, we find almost exactly the same pattern in crime. The more event-driven categories of rhetoric (violent crime, child protection) provide significant Granger causality to crime being named as the nation’s most important problem. Given our previous insight into the ephemeral nature of crime rhetoric, it seems less plausible that public opinion is moved by legislative rhetoric than that legislators act immediately to register their reactions, while it may take several weeks before it registers in a Gallup poll. This does not completely square with the

---

24 The MIP variable is based on all responses coded in PAP data as “International”, “Defense”, or “Trade”. 

33
data, however. The first spike up in attention to crime and child protection occurs in 1999 after
a wave of school shootings, including Columbine. The second peak occurs on October 11, 2000
with final debate on the Victims of Trafficking and Violence Protection Act, after some buildup.
This suggests some substantive opinion leadership on the part of Congress. In any case, it also
suggests that if politicians are pandering in these areas, they are doing so proactively in impressive
anticipation of public opinion.

Conversely, for our time period, public opinion and elite rhetoric on economics appear to be
completely unrelated. One likely explanation is that majorities and minorities move opposite one
another in their desire to discuss the economy. In a good economy, the majority will want to discuss
it; in a bad economy, the minority will want to discuss it. The economy is most salient for the
public when conditions are poor – certainly this is when it is most likely to be cited as the “most
important problem” (Wlezien, 1995) – but responses by the parties are likely to wash out when
aggregated together as they are here. We suspect that when we examine such series separately for
different parties, the results will be different.\footnote{Future work unifying structural, Bayesian vector
autoregression models (Brandt and Freeman 2006) and compositional components could provide further
evidence on cross-issue and eventually cross-party reaction and forecasts. Most importantly, the forecasts
and impulse response functions from these analyses could tell us about the substantive
direction of change across issue areas and time.}

5Discussion

We hope we have demonstrated not only that legislative records hold rich information about
the substance of democratic politics, but that we can extract that information in meaningful ways.
Our data and method offer leverage into the political agenda that supplement and extend prior un-
derstandings in ways that would have been prohibitively expensive or impossible using conventional
methods.

Our next steps are to look in a more systematic way at the micro- and macro-dynamics of the
processes at play here. First, there are strong reasons to believe that the dynamics of different
policy and issue areas are inherently different. Some policy areas are more volatile than others,
some are more responsive to external shocks, some have longer term memories when shocked, and
some have underlying seasonality driven by fiscal years and elections. We seek to understand why
this is the case and how it affects our understanding of political dynamics.

Second, we want to know how these policies interact dynamically. Attention is a finite commod-
ity and its distribution is constrained. More attention to one thing must mean less for something
else. We want to examine how the attention dynamics in one topic area affect those in another.

Third, we want to know what drives these dynamics. At the level of the macro-agenda, when do
we see leadership / manipulation or responsiveness / pandering, and why? How are these affected
by other agendas, like that of the President or as reflected in the media?

Fourth, how does the rhetorical agenda relate to the legislative agenda and actual policy out-
puts? Policy change must be preceded by talk about change, but of course not all talk translates
into action. Can we discern from rhetoric what changes are going to occur before they happen?

These dynamics become even more interesting when we start to disaggregate by party or even
individual. We should, for example, be able to examine the conditions under which majority parties
can successfully constrain the legislative agenda (that we see in roll call votes) to be only the most
advantageous bits of the larger political agenda (that we appear to see in speech). We should also
be able to see whether minority parties engage in heresthetical maneuvers – attempts to shift the
public agenda to issues on which they might become winners – as posited by Riker (1986).

And what drives individual rhetoric? Is there a discernible electoral connection (Mayhew, 1975)
in rhetoric? Is there a reason to expect rhetorical “shirking” among retiring legislators, as has been
investigated in voting patterns (Rothenberg and Sanders, 2000)? We know that parties structure
voting patterns to greater and lesser extents over time and under different institutional conditions;
what are the circumstances in which legislators also talk the party line?

The methods we present here open up great possibilities for such analyses in ways that prior
methods did not. Perhaps most exciting, they also travel beyond English and beyond the Congres-
sional setting, where conventional methods might be prohibitively expensive or impossible to apply.
We hope this might provide an important new window into the nature of democratic politics.
References


Mr. President, I come this morning to again review the lay of the land. As I said a couple of days ago, many of my colleagues, most of our caucus, expressed deep concern alarm, really at the hijacking of the process that went on during the deliberations on the Omnibus appropriations bill. I said at the time, and I believe it ought to be repeated, that I believe the process in the Senate was fair. I have immense respect for the distinguished chairman of the Appropriations Committee. He worked with Members on both sides to accommodate consensus and to reach agreement and the process worked. That process was destroyed at the eleventh hour by some in the administration and by leadership on the Republican side in the House. Changes were demanded. Ultimatums were set. The House and Senate were actually forced to take positions in conference diametrically in opposition to the very positions we took on the Senate floor after a very deliberative debate; positions that I think have great merit.

On an overwhelming vote, the Senate supported the notion that we ought to have country-of-origin labeling. They did it because they believed it is an opportunity for us to enhance our ability to add confidence to consumers' choice, knowing if they buy 100 percent U.S. beef they are not going to buy meat with downer cattle from foreign countries.

Do I understand that the Senate and the House, on both overtime and mad cow, or country of origin, voted by large majorities to have there be a continuation of overtime and to have country-of-origin labeling on all beef that comes into the United States? Did both bodies, by an overwhelming vote, sustain country of origin and elimination of the President's effort to wipe out overtime?

The assistant Democratic leader is correct. That is a succinct summary of what we did. We voted to ensure there be country-of-origin labeling, like 43 other countries have in the world today, knowing we will not be able to export our product to Japan unless it is labeled. We did that.

Figure 1: Example of XML markup of the Congressional Record.
Mr. KOHL. Madam President, I rise today to discuss juvenile crime and juvenile crime prevention programs. We must remember that a strategy to combat juvenile crime consists of a large dose of prevention programs as well as strong enforcement. Juvenile justice programs have proven time and time again that they help prevent crime, strengthen communities, and give children a second chance to succeed and lead healthy lives. It is no secret that robust funding for these programs in the 1990s contributed to a 68 percent drop in juvenile crime from 1994 to 2000. Most importantly, investment in our at-risk children will help prevent a life marred by crime and wasted in prison.

For these programs to succeed, however, they must be priorities for this Congress and for this administration. We fear that we are failing to live up to our responsibility on this essential issue. A little more than 3 months ago, President Bush released his fiscal year 2005 budget proposal. In it, juvenile justice and delinquency programs will receive only about one-third of the funding they received 3 years ago. This is at a time when recent statistics indicate an uptick in juvenile crime and an increase in school murder rates.

We understand that other priorities compete with juvenile justice funding and local crime prevention programs. Yet the amounts we are discussing are so small in the grand scheme of the budget, and the results from the programs so immense, that they mandate our attention.

When the Senate considered the budget resolution, we began to address the shortfalls in juvenile justice funding. I was pleased to work with Senators HATCH and BIDEN on an amendment to restore cuts made to juvenile justice programs and local law enforcement funding. Our amendment represents a step in the right direction by restoring juvenile justice funding to last year’s levels, and reversing the trend of ever-diminishing appropriations for these programs. It is essential that the Kohl-Hatch-Biden amendment that restores juvenile justice funding remain in the final Budget Resolution.

These programs are a wise investment. For every dollar spent on prevention, we save $3 to $4 in costs associated with juvenile crime. Furthermore, law enforcement officials strongly support prevention efforts. A recent poll shows that 71 percent of police chiefs, sheriffs and prosecutors believe that crime prevention efforts would have the greatest impact in reducing youth violence and crime. So for those who may fear that a crime prevention strategy is not “tough” enough on juveniles, we suggest that these programs make sound economic sense and are overwhelmingly endorsed by law enforcement. We must do a better job of funding them.

Let me tell you about two essential programs. In 1992, we established the Title V Local Delinquency Prevention Program. Title V was and remains unique in that it is the only source of federal funding solely dedicated to juvenile crime prevention efforts. More importantly, Title V has proven to be a very successful program that encourages investment, collaboration, and long-range prevention planning by local communities.

Title V programs include preschool and parent training programs, youth mentoring, after-school activities, tutoring, truancy reduction, substance abuse prevention and gang prevention outreach. Through these initiatives, large cities like Milwaukee to small communities like Ladysmith, WI are creating environments that strengthen families and help children avoid crime and develop into productive adults.
Mrs. MURRAY. Mr. President, today I am pleased to submit a resolution designating the week of February 2, 2004, as “National School Counseling Week,” on behalf of my colleagues Senator BIDEN, Senator DORGAN, Senator JOHNSON and Senator DODD. This resolution would honor and celebrate the important work of school counselors, which the Senate has recognized since 1965 through the inclusion of school counseling in the Elementary and Secondary Education Act.

Across the country, there are approximately 95,000 school counselors, including 2,100 in Washington State. School counselors are critical components of a successful school and contribute significantly to the growth and success of students. In fact, school counselors were instrumental in helping students, teachers, and parents deal with the trauma of terrorism on September 11, 2001, and its aftermath. However, despite their important service, counselors are expected to serve, on average 485 students each, and are overwhelmed. The American School Counseling Association, the American Medical Association, and the American Psychological Association recommend the ratio of students to school counselors be 250 students to 1 school counselor.

I want to share just a few examples of how school counselors throughout America are helping students.

In a middle school in southern California, school counselors realized that 257 students were in danger of not passing onto the next grade. They discovered that only 15 percent of the students understood the promotion and retention requirements. The school counselors presented a series of individual and small group lessons on promotion and retention criteria. After the lessons, 100 percent of the students understood the requirements. As a result, 72 of the 257 students, about 28 percent, avoided retention that year.

In a high school in Racine, WI, a math teacher realized that 100 of his students failed algebra in the first quarter of the year. He asked a school counselor for help. Together, they discovered some of the reasons why students were failing. They initiated several programs, such as peer tutoring and homework assistance. As a result, 93 of the 100 students passed algebra by the end of the year and were able to move on to the next level of math.

A school district in Kentucky realized that the retention rate among ninth grade students was unacceptably high. School counselors, teachers and administrators worked together to develop and implement strategies targeted at helping ninth graders move to 10th grade. As a result, retention rates improved in 16 of the 17 high schools in the county in just one year. One school saw the retention rate improve more than 25 percent.

This resolution is merely the beginning of what we need to be doing to support school counselors. We need to reduce the ratio of students to counselors to, at the most, 250 to 1. We need to help schools maintain their funding so that school counselors are not cut from school budgets. And we need to support our school counselors so that they can continue to be integral in the fabric of our schools and help our students achieve success in high school and beyond.
Figure 4. Hierarchical Agglomerative Clustering of $\beta_1, \ldots, \beta_K$. Clustering based on minimizing the maximum Euclidean distance between cluster members. Each cluster is labeled with a topic name, followed by the percentage of documents and words, respectively, in that cluster.
Figure 5: Partial crosswalk comparison of rhetorical clustering with other policy classification schemes. Filled cells indicate the existence of some shared topical content.
Figure 6: The Probability of a Randomly Chosen Document being on ‘Judicial Nominations’ ($\pi_{\text{judicial}}$). Values of $\pi$ that are farther than 7 days from a day in which legislative speech occurred have been whited out.
Figure 7: The Probability of a Randomly Chosen Document being on ‘Supreme Court / Constitutional’ ($\pi_{judicial2}$). Values of $\pi$ that are farther than 7 days from a day in which legislative speech occurred have been whited out.
Figure 8: The Probability of a Randomly Chosen Document being on 'Abortion' ($\pi_{\text{abortion}}$). Values of $\pi$ that are farther than 7 days from a day in which legislative speech occurred have been whited out.
Figure 9: Estimated persistence of shocks to selected topics, AR-7 model.
Figure 10: Persistence and volatility in the daily dynamics of topic attention. Both derived from independent models of daily word/topic logratios (as given by topic detection model), AR-7 with monthly dummies.
Figure 11: The Probability of a Randomly Chosen Document being on ‘Defense [Use of Force]’, ‘International [Arms]’, or ‘International [Diplomacy / Human Rights]’. ‘Use of Force’ is at the bottom in dark brown, ‘Arms’ is in the middle in light blue, and ‘Diplomacy / Human Rights’ is at the top in yellow. Values of $\pi$ that are farther than 7 days from a day in which legislative speech occurred have been whited out.
Figure 12: The Probability of a Randomly Chosen Document being on ‘Symbolic [Military]’ or ‘Defense [Use of Force]’. Symbolic is at the bottom in brown and Force is above in light blue. Values of $\pi$ that are farther than 7 days from a day in which legislative speech occurred have been whited out.
Figure 13: Monthly Most Important Problem = Defense / International (log-odds) and monthly logratios of words/topic in selected categories. These are typical inputs for the Granger causality test.
Figure 14: Monthly Most Important Problem = Crime (log-odds) and monthly logratios of words/topic in selected categories. These are typical inputs for the Granger causality test.
<table>
<thead>
<tr>
<th>Topic Labels</th>
<th>%a</th>
<th>Clarifying Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Judicial Nominations</td>
<td>1.0/2.4</td>
<td></td>
</tr>
<tr>
<td>2. Supreme Court / Constitutional</td>
<td>1.1/3.0</td>
<td>incl. impeachment, DOJ, marriage, flag-burning</td>
</tr>
<tr>
<td>3. Campaign Finance</td>
<td>0.9/2.4</td>
<td></td>
</tr>
<tr>
<td>4. Abortion</td>
<td>0.5/1.1</td>
<td></td>
</tr>
<tr>
<td>5. Law &amp; Crime 1 [Violence/Drugs]</td>
<td>1.3/1.8</td>
<td>violence, drug trafficking, police, prison</td>
</tr>
<tr>
<td>6. Child Protection</td>
<td>0.9/2.6</td>
<td>tobacco, alcohol, drug abuse, school violence, abuse</td>
</tr>
<tr>
<td>7. Health 1 [Medical]</td>
<td>1.5/2.4</td>
<td>emph. disease, prevention, research, regulation</td>
</tr>
<tr>
<td>8. Social Welfare</td>
<td>2.0/2.8</td>
<td></td>
</tr>
<tr>
<td>9. Education</td>
<td>1.8/4.6</td>
<td></td>
</tr>
<tr>
<td>10. Armed Forces 1 [Manpower]</td>
<td>1.0/1.5</td>
<td>incl. veterans’ issues</td>
</tr>
<tr>
<td>11. Armed Forces 2 [Infrastructure]</td>
<td>2.3/3.0</td>
<td>incl. bases and civil defense</td>
</tr>
<tr>
<td>12. Intelligence</td>
<td>1.4/3.9</td>
<td>incl. terrorism and homeland security</td>
</tr>
<tr>
<td>13. Crime 2 [Federal]</td>
<td>1.8/2.7</td>
<td>incl. the FBI, immigration, white collar crime</td>
</tr>
<tr>
<td>14. Environment 1 [Public Lands]</td>
<td>2.2/2.5</td>
<td>incl. water management, resources, Native Americans</td>
</tr>
<tr>
<td>15. Commercial Infrastructure</td>
<td>2.0/2.9</td>
<td>incl. transportation and telecom</td>
</tr>
<tr>
<td>16. Banking and Finance</td>
<td>1.1/3.1</td>
<td>incl. corporations, small business, torts, bankruptcy</td>
</tr>
<tr>
<td>17. Labor 1 [Workers, esp Retirement]</td>
<td>1.0/1.5</td>
<td>emph. conditions and benefits, esp. pensions</td>
</tr>
<tr>
<td>18. Debt / Deficit / Social Security</td>
<td>1.7/4.6</td>
<td></td>
</tr>
<tr>
<td>19. Labor 2 [Employment]</td>
<td>1.4/4.5</td>
<td>incl. jobs, wages, general state of the economy</td>
</tr>
<tr>
<td>20. Taxes</td>
<td>1.1/2.7</td>
<td>emph. individual taxation, incl. income and estate</td>
</tr>
<tr>
<td>21. Energy</td>
<td>1.4/3.3</td>
<td>incl. energy supply and prices, environmental effects</td>
</tr>
<tr>
<td>22. Environment 2 [Regulation]</td>
<td>1.1/2.8</td>
<td>incl. pollution, wildlife protection</td>
</tr>
<tr>
<td>23. Agriculture</td>
<td>1.2/2.5</td>
<td></td>
</tr>
<tr>
<td>24. Foreign Trade</td>
<td>1.1/2.4</td>
<td></td>
</tr>
<tr>
<td>25. Procedural 3 [Legislation 1]</td>
<td>2.0/2.8</td>
<td></td>
</tr>
<tr>
<td>27. Health 2 [Economics - Seniors]</td>
<td>1.0/2.6</td>
<td>incl. Medicare and prescription drug coverage</td>
</tr>
<tr>
<td>28. Health 3 [Economics - General]</td>
<td>0.8/2.3</td>
<td>incl. provision, access, costs</td>
</tr>
<tr>
<td>29. Defense [Use of Force]</td>
<td>1.4/3.7</td>
<td>incl. wars/interventions, Iraq, Bosnia, etc.</td>
</tr>
<tr>
<td>30. International Affairs [Diplomacy]</td>
<td>1.9/3.0</td>
<td>incl. human rights, organizations, China, Israel, etc.</td>
</tr>
<tr>
<td>31. International Affairs [Arms Control]</td>
<td>0.9/2.3</td>
<td>incl. treaties, nonproliferation, WMDs</td>
</tr>
<tr>
<td>32. Symbolic [Tribute - Living]</td>
<td>1.9/1.3</td>
<td></td>
</tr>
<tr>
<td>33. Symbolic [Tribute - Constituent]</td>
<td>3.2/1.9</td>
<td></td>
</tr>
<tr>
<td>34. Symbolic [Remembrance - Military]</td>
<td>2.3/1.9</td>
<td>incl. tributes to other public servants, WWII Memorial</td>
</tr>
<tr>
<td>35. Symbolic [Remembrance - Nonmilitary]</td>
<td>2.4/2.3</td>
<td></td>
</tr>
<tr>
<td>36. Symbolic [Congratulations - Sports]</td>
<td>0.6/0.4</td>
<td></td>
</tr>
<tr>
<td>37. Jesse Helms re Debt</td>
<td>0.5/0.1</td>
<td>almost daily deficit / debt ‘boxscore’ speeches</td>
</tr>
<tr>
<td>38. Gordon Smith re Hate Crime</td>
<td>0.4/0.1</td>
<td>almost daily speeches on hate crime</td>
</tr>
<tr>
<td>39. Procedural 1 [Housekeeping 1]</td>
<td>20.4/1.5</td>
<td></td>
</tr>
<tr>
<td>40. Procedural 5 [Housekeeping 3]</td>
<td>15.5/1.0</td>
<td></td>
</tr>
<tr>
<td>41. Procedural 6 [Housekeeping 4]</td>
<td>6.5/1.6</td>
<td></td>
</tr>
<tr>
<td>42. Procedural 2 [Housekeeping 2]</td>
<td>2.4/0.8</td>
<td></td>
</tr>
</tbody>
</table>

*aPercentage of documents and words in each topic.

Table 1: Topic labels and descriptive statistics for 42-topic model.
<table>
<thead>
<tr>
<th>Topic (Short Label)</th>
<th>Keys</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Judicial Nominations</td>
<td>nomine, confirm, nomin, circuit, hear, court, judg, judici, case, vacanc</td>
</tr>
<tr>
<td>2. Constitutional</td>
<td>case, court, attornei, supreme, justic, nomin, judg, m, decis, constitut</td>
</tr>
<tr>
<td>3. Campaign Finance</td>
<td>campaign, candid, elect, monei, contribut, polit, soft, ad, parti, limit</td>
</tr>
<tr>
<td>4. Abortion</td>
<td>procedur, abort, babi, thi, life, doctor, human, ban, decis, or</td>
</tr>
<tr>
<td>5. Crime 1 [Violent]</td>
<td>enforce, act, crime, gun, law, victim, violenc, abus, prevent, juvenil</td>
</tr>
<tr>
<td>6. Child Protection</td>
<td>gun, tobacco, smoke, kid, show, firearm, crime, kill, law, school</td>
</tr>
<tr>
<td>7. Health 1 [Medical]</td>
<td>diseases, cancer, research, health, prevent, patient, treatment, devic, food</td>
</tr>
<tr>
<td>8. Social Welfare</td>
<td>care, health, act, home, hospit, support, children, educ, student, nurs</td>
</tr>
<tr>
<td>9. Education</td>
<td>school, teacher, educ, student, children, test, local, learn, district, class</td>
</tr>
<tr>
<td>10. Military 1 [Manpower]</td>
<td>appropri, va, forc, militari, care, reserv, serv, men, guard, member</td>
</tr>
<tr>
<td>11. Military 2 [Infrastructure]</td>
<td>intellig, homeland, commiss, depart, agenc, director, secur, base, defens</td>
</tr>
<tr>
<td>12. Intelligence</td>
<td>act, inform, enforce, record, law, court, section, crimin, internet, investig</td>
</tr>
<tr>
<td>13. Crime 2 [Federal]</td>
<td>land, water, park, act, river, natur, wildlif, area, conserv, forest</td>
</tr>
<tr>
<td>14. Environment 1 [Public Lands]</td>
<td>small, busi, act, highwai, transport, internet, loan, credit, local, capit</td>
</tr>
<tr>
<td>15. Commercial Infrastructure</td>
<td>bankruptci, bank, credit, case, ir, compani, file, card, financi, lawyer</td>
</tr>
<tr>
<td>16. Labor 1 [Workers]</td>
<td>worker, social, retir, benefit, plan, act, employ, pension, small, employe</td>
</tr>
<tr>
<td>17. Debt / Social Security</td>
<td>social, year, cut, budget, debt, spend, balanc, deficit, over, trust</td>
</tr>
<tr>
<td>18. Labor 2 [Employment]</td>
<td>job, worker, pai, wage, economi, hour, compani, minimum, overtim</td>
</tr>
<tr>
<td>19. Taxes</td>
<td>act, again, tax, cut, incom, pai, estat, over, relief, marriag, than, penalti</td>
</tr>
<tr>
<td>20. Energy</td>
<td>energi, fuel, ga, oil, price, produce, electr, renew, natur, suppli</td>
</tr>
<tr>
<td>21. Agriculture</td>
<td>wast, land, water, site, forest, nuclear, fire, mine, environment, road</td>
</tr>
<tr>
<td>22. Environment 2 [Regulation]</td>
<td>farmer, price, produc, farm, crop, agricultur, disast, compact, food, market</td>
</tr>
<tr>
<td>23. Trade</td>
<td>trade, agreement, china, negoti, import, countri, worker, unit, world, free</td>
</tr>
<tr>
<td>24. Procedural 3</td>
<td>mr, consent, unanim, order, move, senat, ask, amend, presid, quorum</td>
</tr>
<tr>
<td>25. Procedural 4</td>
<td>leader, major, am, senat, move, issu, hope, week, done, to</td>
</tr>
<tr>
<td>26. Health 2 [Seniors]</td>
<td>senior, drug, prescript, medicar, coverag, benefit, plan, price, beneficiari</td>
</tr>
<tr>
<td>27. Health 3 [Economics]</td>
<td>patient, care, doctor, health, insur, medic, plan, coverag, decis, right</td>
</tr>
<tr>
<td>29. International [Diplomacy]</td>
<td>unit, human, peac, nato, china, forc, intern, democraci, resolat, europ</td>
</tr>
<tr>
<td>30. International [Arms]</td>
<td>test, treati, weapon, russia, nuclear, defens, unit, missil, chemic</td>
</tr>
<tr>
<td>31. International [Living]</td>
<td>serv, hi, career, dedic, john, posit, honor, nomin, dure, miss</td>
</tr>
<tr>
<td>32. Symbolic [Living]</td>
<td>recog, dedic, honor, serv, insert, contribut, celebr, congratul, career</td>
</tr>
<tr>
<td>33. Symbolic [Constituent]</td>
<td>honor, men, sacrifi, menori, dedic, freedom, di, kill, serve, soldier</td>
</tr>
<tr>
<td>34. Symbolic [Military]</td>
<td>great, hi, paul, john, alwai, reagan, him, serv, love</td>
</tr>
<tr>
<td>35. Symbolic [Nonmilitary]</td>
<td>team, game, plai, player, win, fan, basebal, congratul, record, victori</td>
</tr>
<tr>
<td>36. Symbolic [Sports]</td>
<td>hundr, at, four, three, ago, of, year, five, two, the</td>
</tr>
<tr>
<td>37. J. Helms re Debt</td>
<td>of, and, in, chang, by, to, a, act, with, the, hate</td>
</tr>
<tr>
<td>38. G. Smith re Hate Crime</td>
<td>order, without, the, from, object, recog, so, second, call, clerk</td>
</tr>
<tr>
<td>39. Procedural 1</td>
<td>consent, unanim, the, of, mr, to, order, further, and, consider</td>
</tr>
<tr>
<td>40. Procedural 5</td>
<td>mr, consent, unanim, of, to, at, order, the, consider, follow</td>
</tr>
<tr>
<td>41. Procedural 6</td>
<td>of, mr, consent, unanim, and, at, meet, on, the, am</td>
</tr>
<tr>
<td>42. Procedural 2</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: For each topic, the top ten (or so) key stems that best distinguish the topic from all others. Keywords have been sorted here by rank(\(\beta_k w\)) + rank(\(\tau_k w\)), as defined in the text. Lists of the top forty keywords for each topic and related information are provided in the web appendix (?).
Key Words for “Judicial Nominations” Topic


nomine, confirm, nomin, circuit, hear, court, judg, judici, case, vacanc, judiciari, major, clinton, appeal, record, been, bush, last, white, ha, ninth, appoint, hold, suprem, bench, republican, consent, were, than, month, consid, committe, of, filibust, justic, dure, never, posit, democrat, view


consensu, progress, chosen, ideolog, utah, lifetim, moder, pack, father, divers, vacanc, lowest, pennsylvania, insist, scalia, divid, sutton, brown, florida, federalist, privaci, 1995, term, georg, far, violat, bybe, discrimin, prado, white, block, leadership, fourth, refus, abus, 100, check, activ, agenda, instead


clerk, up-or-down, supermajor, oppon, myer, awar, abort, cum, harvard, leahi, qualif, newspap, confidenc, tactic, prepar, lack, solicitor, emery, top, cost, tonight, conclud, memoranda, prior, columbia, janic, perform, present, face, bork, shall, review, miguel, tax, young, dure, parent, rank, especi, import

Table 3: Partisan keys are the terms that, given the classification of a 108th Senate speech as “Judical Nominations” by the method given here, most strongly indicate that the speaker belongs to the particular party. These are calculated using the “rhetorical ideal point estimation” method described by Monroe and Maeda (2004).
<table>
<thead>
<tr>
<th>Category</th>
<th>9/12/01</th>
<th>Post-9/11</th>
<th>9/12/01</th>
<th>Post-9/11</th>
<th>Bush II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judicial Nominations</td>
<td>5.28</td>
<td>3.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constitutional</td>
<td>-2.19</td>
<td></td>
<td>-2.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Campaign Finance</td>
<td>-3.19</td>
<td></td>
<td></td>
<td></td>
<td>-2.25</td>
</tr>
<tr>
<td>Abortion</td>
<td>-2.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crime 1 [Violent]</td>
<td></td>
<td></td>
<td>-6.68</td>
<td></td>
<td>-3.01</td>
</tr>
<tr>
<td>Child Protection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health 1 [Medical]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Welfare</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>-3.87</td>
<td>-3.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armed Forces 1 [Manpower]</td>
<td>2.83</td>
<td></td>
<td>2.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armed Forces 2 [Infrastructure]</td>
<td>2.19</td>
<td>1.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intelligence</td>
<td>5.00</td>
<td>3.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crime 2 [Federal]</td>
<td></td>
<td></td>
<td>2.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environment 1 [Public Lands]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercial Infrastructure</td>
<td>-2.79</td>
<td>-2.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banking / Finance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor 1 [Workers]</td>
<td>2.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Debt / Social Security</td>
<td>-3.60</td>
<td>-2.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labor 2 [Employment]</td>
<td>4.06</td>
<td>3.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taxes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy</td>
<td>-2.30</td>
<td>3.96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture</td>
<td>-3.23</td>
<td>-2.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Procedural 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Procedural 4</td>
<td>3.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health 2 [Seniors]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health 3 [Economics]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defense [Use of Force]</td>
<td>6.61</td>
<td>6.07</td>
<td>-2.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>International Affairs [Nonmilitary]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>International Affairs [Arms]</td>
<td>2.17</td>
<td>-4.62</td>
<td>2.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symbolic [Living]</td>
<td>3.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symbolic [Constituent]</td>
<td>3.88</td>
<td>2.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symbolic [Military]</td>
<td>2.02</td>
<td>6.38</td>
<td>2.02</td>
<td>3.41</td>
<td></td>
</tr>
<tr>
<td>Symbolic [Nonmilitary]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symbolic [Sports]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jesse Helms re Debt</td>
<td>-9.56</td>
<td>-4.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gordon Smith re Hate Crime</td>
<td>-4.56</td>
<td>10.06</td>
<td>-3.79</td>
<td>3.59</td>
<td>7.40</td>
</tr>
<tr>
<td>Procedural 1</td>
<td>2.22</td>
<td>2.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Procedural 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Procedural 5</td>
<td>2.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Procedural 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Procedural 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Each series (centered logratios of daily words) modelled as an independent AR-7 with monthly dummies. The AR and monthly coefficients are not reported here. Values shown are \( t \)-values, such that \( |t| > 1.96 \).
<table>
<thead>
<tr>
<th>Category</th>
<th>Speech → MIP</th>
<th>MIP → Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Defense</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defense [Use of Force]</td>
<td>0.048</td>
<td>0.577</td>
</tr>
<tr>
<td>Int’l Affairs [Nonmilitary]</td>
<td>0.201</td>
<td>0.307</td>
</tr>
<tr>
<td>Int’l Affairs [Arms]</td>
<td>0.498</td>
<td>0.018</td>
</tr>
<tr>
<td>Armed Forces 1 [Manpower]</td>
<td>0.029</td>
<td>0.514</td>
</tr>
<tr>
<td>Armed Forces 2 [Infrastructure]</td>
<td>0.047</td>
<td>0.984</td>
</tr>
<tr>
<td>Intelligence</td>
<td>0.039</td>
<td>0.137</td>
</tr>
<tr>
<td>Trade</td>
<td>0.929</td>
<td>0.075</td>
</tr>
<tr>
<td><strong>Crime</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Law / Crime 1 [Violent]</td>
<td>0.002</td>
<td>0.835</td>
</tr>
<tr>
<td>Child Protection</td>
<td>0.022</td>
<td>0.895</td>
</tr>
<tr>
<td>Crime 2 [Federal]</td>
<td>0.092</td>
<td>0.181</td>
</tr>
<tr>
<td><strong>Economics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Debt / Social Security</td>
<td>0.957</td>
<td>0.602</td>
</tr>
<tr>
<td>Labor 2 [Employment]</td>
<td>0.927</td>
<td>0.519</td>
</tr>
<tr>
<td>Taxes</td>
<td>0.166</td>
<td>0.994</td>
</tr>
<tr>
<td>Banking / Finance</td>
<td>0.635</td>
<td>0.356</td>
</tr>
<tr>
<td>Labor 1 [Workers]</td>
<td>0.658</td>
<td>0.107</td>
</tr>
</tbody>
</table>

Table 5: Bivariate Granger causality tests (F-tests) for ‘Most Important Problem’ Gallup responses and legislative speech in related categories. Entries are p-values under the null hypothesis that the relationship in question does not hold.
A Estimation

We use the Expectation Conditional Maximization (ECM) algorithm to fit this model. Schematically, the algorithm proceeds as follows.

1. Set initial values $\theta^{(0)}, \omega^{(0)}, \eta^{(0)}, V_t^{(0)}$ and $W_t^{(0)}$.
2. $i = 0$
3. do until convergence
   a. Calculate $\hat{z}_{dk} = E(Z_{dk} | Y, \theta^{(i)}, \omega^{(i)}, \eta^{(i)}, V_t^{(i)}, W_t^{(i)})$ for $d = 1, \ldots, D$ and $k = 1, \ldots, K$
   b. Set $\omega^{(i+1)} = \arg\max_\omega p(\omega | \hat{Z}, Y, \theta^{(i)}, \eta^{(i)}, V_t^{(i)}, W_t^{(i)})$
   c. Set $\eta^{(i+1)} = \arg\max_\eta p(\eta | \hat{Z}, Y, \theta^{(i)}, \omega^{(i+1)}, V_t^{(i)}, W_t^{(i)})$
   d. Set $V_t^{(i+1)} = \arg\max_{V_t} p(V_t | \hat{Z}, Y, \theta^{(i)}, \omega^{(i+1)}, \eta^{(i+1)}, V_t^{(i)}, W_t^{(i)})$ for $t = 1, \ldots, T$
   e. Set $W_t^{(i+1)} = \arg\max_{W_t} p(W_t | \hat{Z}, Y, \theta^{(i)}, \omega^{(i+1)}, \eta^{(i+1)}, V_{t-1}^{(i)}, W_t^{(i)})$ for $t = 1, \ldots, T$
   f. Set $\theta^{(i+1)} = \arg\max_\theta p(\theta | \hat{Z}, Y, \omega^{(i+1)}, \eta^{(i+1)}, V_t^{(i+1)}, W_t^{(i+1)})$
   g. $i = i + 1$

More specifically, each of the steps above is accomplished as follows.

A.1 Step 3a (Conditional Expectation of $Z$)

Standard calculations (McLachlan and Peel, 2000) reveal that

$$\hat{z}_{dk} = \frac{\pi_{s(d)k} \prod_{w=1}^W \theta_{kw}^{y_{dw}}}{\sum_{j=1}^K \pi_{s(d)j} \prod_{w=1}^W \theta_{jw}^{y_{dw}}}.$$ 

A.2 Step 3b (Conditional Maximization of $\omega$)

A simple analytic solution for this step is not available. Instead, we use the BFGS algorithm to numerically maximize

$$p(\omega_t | \hat{Z}, Y, \theta^{(i)}, \eta^{(i)}, V_t^{(i)}, W_t^{(i)}) \propto f_N(\omega_t | F_t \eta_t, V_t) \prod_{d,s(d)=t} f_M(z_d | 1, \pi_t)$$

where $f_N(\cdot | \mu, \Sigma)$ is the normal density with mean $\mu$ and variance-covariance matrix $\Sigma$; $f_M(\cdot | n, \rho)$ is the multinomial mass function with sample size $n$ and probability vector $\rho$; and

$$\pi_t = \left( \frac{\exp(\omega_{t1})}{1 + \sum_{j=1}^{K-1} \exp(\omega_{tj})}, \ldots, \frac{\exp(\omega_{t(K-1)})}{1 + \sum_{j=1}^{K-1} \exp(\omega_{tj})}, \frac{1}{1 + \sum_{j=1}^{K-1} \exp(\omega_{tj})} \right).$$

Handling each of the $\omega_t$ vectors separately is appropriate since they are conditionally independent.
A.3 Step 3c (Conditional Maximization of $\eta$)

Because of the conditionally Gaussian structure of this portion of the model we can use standard results for Gaussian dynamic linear models (West and Harrison, 1997) to calculate the value of $\eta$ that maximizes the posterior conditional on the other model parameters. Specifically:

1. for $t = 1, \ldots, T$ {
   (a) $a_t = G_t m_{t-1}$
   (b) $R_t = G_t C_{t-1} G_t' + W_t$
   (c) $f_t = F_t' a_t$
   (d) $Q_t = F_t' R_t F_t + V_t$
   (e) $A_t = R_t F_t Q_t^{-1}$
   (f) $e_t = \omega_t - f_t$
   (g) $m_t = a_t + A_t e_t$
   (h) $C_t = R_t - A_t Q_t A_t'$

2. $\eta_T = m_T$

3. for $t = (T - 1), \ldots, 1$ {
   (a) $B_t = C_t G_{t+1} R_{t+1}^{-1}$
   (b) $\eta_t = m_t + B_t (\eta_{t+1} - a_{t+1})$

Note that $f_t$ is the one-step-ahead forecast of $\omega_t$. If we are interested in forecasting this quantity can also be easily stored.

A.4 Step 3d (Conditional Maximization of $V$)

Conditional on the other model parameters, each diagonal element $V_{ii}$ of $V$ follows an inverse gamma distribution with mode at

$$\frac{b_0 + \sum_{t \in T} (\omega_{ti} - [F_t' \eta_t]_i)^2}{(a_0 + |T|)/2 + 1}$$

where $T$ denotes the set of days Congress was in session and $[F_t' \eta_t]_i$ is the $i$th element of $F_t' \eta_t$.

A.5 Step 3e (Conditional Maximization of $W$)

Conditional on the other model parameters, each diagonal element $W_{ii}$ of $W$ follows an inverse gamma distribution with mode at

$$\frac{d_0 + \sum_{t=2}^{T} (\eta_{ti} - [G_t \eta_{t-1}]_i)^2}{(c_0 + T - 1)/2 + 1}$$

where $[G_t \eta_{t-1}]_i$ is the $i$th element of $G_t \eta_{t-1}$.

A.6 Step 3f (Conditional Maximization of $\theta$)

Simple calculations show that the value of $\theta$ that maximizes the conditional posterior has typical element

$$\theta_{kw} = \frac{\sum_{d=1}^{D} \hat{z}_{dk} y_{dw} + \lambda_{kw}}{\sum_{j=1}^{W} \left\{ \sum_{d=1}^{D} \hat{z}_{dk} y_{dj} + \lambda_{kj} \right\}}$$