Political Dialogue and the Great Recession: How Democratic Institutions Respond to Crises

Abstract

When faced with crises or newly emergent issues, policymakers are often forced to address problems with which they have little experience. In these scenarios, uncertainty about problem definition and ambiguity regarding comparisons between the various attributes of the problem frequently dominate debates among policymakers. In this paper, we clarify the concept of uncertainty as used in policy debates and develop and validate a topic model-based measurement approach. The approach is designed to trace variation in speaker uncertainty over time as expressed in text data. We then apply this approach to a novel dataset of Congressional hearing transcripts on financial policy held between 2002-2011. Generally, we find that financial policy discussions were characterized by increased uncertainty following the onset of the crisis, followed by a decline as the crisis subsided.

1 Introduction

How do open, democratic governments process policy-relevant information during crisis situations? And, how do patterns of information processing during crises differ from those observed under more conventional conditions? Severe crises can cause a breakdown in prevailing worldviews, offering opportunities to institute major policy changes but leaving legislators in need of information regarding the challenges they face. In these situations, legislators place a premium on information that allows them to understand the problem at hand, and the connections between the problem and proposed solutions.

Generally, policy-relevant information comes "packaged" in the form of (1) assertions about the nature of problem, and (2) policy arguments about the appropriate actions to take. During a crisis, uncertainty about the nature of the problem and the correct course of action increases. However, this uncertainty takes a particular form. Rather than uncertainty regarding the weights of relevant issue dimensions, decision-makers often face deep ambivalence or even loss of confidence in their basic understanding of the problem. Previous frames become irrelevant, forcing policymakers to re-examine the problem from a new perspective. As a consequence, arguments become more complex and uncertain as decision-makers struggle to impose a new understanding onto conversations surrounding the issue.

In this paper, we outline a measurement approach intended to detect the kinds of patterns described above. In particular, we develop a novel language-based measure of individual-level uncertainty that assesses the complexity and topical diversity of a given policy conversation. Text-based approaches frequently have difficulty tracking temporal dynamics, particularly when the dependent variable of interest is content-based (e.g. attention to a particular topic or policy area) [Lowe and Benoit, 2013]. However, our approach is less susceptible to these problems. Uncertainty in a political debate, we suggest, is most closely related to the ways attributes structure (or frame) an issue, rather than the details of the attributes under consideration [Chong and Druckman, 2007, Jones, 1994]. As we argue, our dependent variable of interest is therefore the dimensionality of the conversation itself, rather than the attention paid to any particular problem dimension.

We apply this approach to study policy debates in the US Congress before, during, and after the financial crisis and subsequent Great Recession (2007-2008). Throughout this process, we rely on an original, full-text dataset of Congressional hearings held from 2002-2011. Generally, we find that hearing participants discussed a wider range of ideas following the onset of the financial crisis than in the years preceding preceding the event. This conversational complexity declined as the crisis abated, and as Congress enacted new laws and provisions addressing the problem at hand.

2 Uncertainty and Complexity in Policy Debates

2.1 Setting the Scene

On April 2, 2009, Chairman Edolphus Towns of the House Committee on Oversight and Government Reform initiated a hearing dedicated to investigating the crisis:

> [O]ur financial regulatory system has failed the American people. Dangerous and unacceptable levels of risk were allowed to build up in our financial system, leading to the catastrophic failure of our Nation's economy. But exactly why this happened and who was responsible remains unclear. [...] The American people deserve clear answers to how and why this failure occurred, who was responsible and who benefited, why the existing regulatory system failed, and what steps need to be taken going forward in this regard.[1]

The committee then heard testimony from Maurice Greenberg, the former chairman of American International Group (AIG) – the largest single recipient of federal bailout money – regarding AIG's role in the collapse. Greenberg's testimony and the accompanying discussion covered a wide array of possible contributors to the crisis, including unregulated

[1] Edolphus Towns, U.S. Congress, House, Oversight Committee, *AIG Collapse and Federal Rescue*, 111th Cong., 1st Sess., April 2, 2009.

credit-default swaps, subprime mortgages, management practices in financial institutions, executive compensation, the repeal of Glass-Stegall, and unregulated security lending. The committee's discussion of possible solutions to the financial crisis was no less diverse. Some of the stop-gap solutions they proposed included reinstating Glass-Stegall, increasing bank capital requirements, and forming a bi-partisan commission that would further investigate the causes of the crisis. The number and diversity of factors and actors raised by the committee led Chairman Towns to conclude that the roots of the meltdown appeared to be a "conglomeration of problems that brought us to this point of economic crisis." The chairman closed the hearing with an assertion that they would continue "to try to get to the bottom of this":

> And before we adjourn, let me state that this committee intends to continue its examination of the financial crisis until we get a much better understanding of what caused it. As the old saying goes, the past is prologue. Until we can explain what went wrong, how can we chart the best course for reform?[2]

This hearing illustrates several aspects of decision-making in Congress. There are sound theoretical reasons to expect that as uncertainty rises during a crisis, the *complexity* of the policy space - which we define as the diversity of the arguments, assertions, and causal claims made in legislative testimony - increases, only to decrease as the crisis abates. As we assert in subsequent discussion, the complexity of argumentation in a policy discussion generally indicates the extent of uncertainty in policymaking; so, as complexity increases, so too does uncertainty. As the crisis abates, policymakers settle on a dominant frame for understanding the problem, and uncertainty (and complexity of discussion) decline.

While this idea may seem simple enough, its implications are far-reaching. Many current explanations of lawmaking center on preferences and ideologies, with little room for information-processing and problem-solving [Krehbiel, 1999, Poole and Rosenthal, 2000, McCarty et al., 2013]. Preferences in politics generally center on policy solutions, not problem

[2] *Ibid.*

identification. Yet if our central thesis is correct, and break-downs in the understanding of the relationship between policy problems and the solutions adopted to address those problems occur, then the simple preference model fails to capture important aspects of the policymaking process.

2.2 Two Methods of Search

One fundamental insight of the policy studies literature is that the definition and prioritization of problems works by a different set of mechanisms than the process of selecting an appropriate solution. In a series of experiments, Newell and Simon [1972] found that subjects generally failed to conduct detailed examinations of the problem space, and worked through solutions heuristically. Cohen et al. [1972] and Kingdon [1984] demonstrate the same dynamic in organizations and the policymaking processes.

More recent work indicates that policymaking organizations engage in a narrow and focused *expert search* when issues are well-understood and information is focused. This pattern can usually be observed in the solution space, after a problem is defined (or treated as defined by participants). When issues are messy and multidimensional, as is often the case of a new issue, complex problem area, or during a crisis, policymakers instead tend to focus on information prioritization and problem definition. Information prioritization requires policymakers to filter through the oversupply of information often characterized by divergent perspectives - a process sometimes termed *entropic search* [Baumgartner and Jones, 2015].

While these two strategies operate in two different contexts, no policy will permanently reside in either the problem or solution space. Shocks or crises as well as new information can lead to policies requiring redefinition [Baumgartner et al., 2009]. As problem definitions collapse, policymakers are best served when they adjust their information gathering strategies from one of narrow, expert-driven search to broad information prioritization from a broad range of suppliers. Actors seeking to define the problem to mirror their preferences compete to supply information to policymakers. This process offers legislators the opportunity to examine a problem from a wide range of perspectives, helping to lower uncertainty about the dimensionality of a problem. As a policy becomes well-defined and moves into the solution space, the policymaking process returns to the expert-search pattern.

2.3 Cognitive Foundations

For the purposes of this paper, then, our fundamental phenomenon of interest is the manifestation of uncertainty in problems of individual choice. Here, we use the term *uncertainty* to refer to issue spaces characterized by high levels of diversity in the topics under consideration. High-uncertainty problems compel relevant actors to simultaneously hold both positive and negative ideas about an object, which results in cognitive tension and difficulties with decision-making[Cacioppo and Berntson, 1994]. Individuals operating in high-uncertainty situations are therefore likely to discuss a greater number of aspects of an underlying problem, or divide their attention more evenly between a given set of issues. At a group level, uncertainty can manifest as a collection of internally conflicted individuals—each of whom individually engages with a large number of problem dimensions—or as a group whose members are internally clear regarding the appropriate problem definition, but collectively disagree [Albertson et al., 2005]. In the context of Congressional decision-making following the financial crisis, when multiple problem dimensions or policy solutions are under consideration, lawmakers may appear ambivalent regarding their preferred theory, as they flit from one topic to another [Albertson et al., 2005, Alvarez and Brehm, 1995].

More generally, choices regarding a policy problem of interest are framed by individuals' understanding of these issue attributes, which are weighted according to that individual's assessment of their relative importance. As we use the term in this paper, an *attribute* is a quality, character, or inherent part of an object or situation. Attributes may take the form of facts, feelings, theories, or arguments about the problem or situation in question. The relevant attributes associated with familiar tasks are generally well-understood (e.g. price, color, or year of a used car), but many choice situations do not posses such clear-cut decision standards. We refer to the cognitive challenges experienced by decision-makers in the latter context as *attribute uncertainty.*

Decision-makers face attribute uncertainty when they are unsure of the appropriate decision frame. In these situations, individuals are unsure of which attributes are relevant to the choice, and the proper weights to attach to those dimensions. Attribute uncertainty captures confusion and ambivalence about the understanding, characterization, or framing of the problem-space.[3] How individuals understand an issue – that is, the weights they attribute to each seemingly relevant dimension – impacts the ways they choose to address that issue. Unfortunately, these attribute weights are not directly observable; however, we can obtain some information about the weights individuals place on various problem components by examining the ideas they discuss. Discussion of a greater diversity of attributes indicates greater uncertainty regarding the relative importance of the ideas under consideration.

2.4 Attribute Uncertainty and Dimensionality Reduction in Policy Choice

Studies of legislative policy-making processes have largely focused on voting and other processes that can be well-summarized using low-dimensional data representations. Though this body of work is clearly important, it only examines a part of the policy-making process, and misses important variation in problem definition and solution search that occur prior to the final voting stage. Under ordinary circumstances, ideology, parties, institutional structures, and other devices tend to reduce uncertainty, giving participants firm cues regarding the appropriate understanding of an issue [Kingdon, 1984]. In these situations, ideology can usefully be represented and interpreted in a low- or single-dimensional space Hinich and Munger [1994]. The utility of this approach is frequently substantial, and empirically ob-

[3]Ambivalence exists when decision-makers know the attributes that structure a choice, but are conflicted about how to balance or trade-off the attributes [Alvarez and Brehm, 1995].

servable; Poole and Rosenthal [2000], for example, famously conclude that low-dimensional solutions can account for most of the variation in Congressional voting patterns through US history.

Under other circumstances, however, these dimensionality reduction mechanisms are likely to be less influential. In particular, we argue that uncertainty is most likely to rise (1) when issues that reach the public agenda are new and ill-understood or (2) when a crisis or other external event forces politicians to re-define an issue by recognizing one or more attributes that were previously unknown.[4] This uncertainty may abate (and the dimensionality of the choice space may decrease) as the new issue becomes more understandable or the crisis eases [Jones and Baumgartner, 2005].

Importantly, the kinds of situations in which we might expect to see increased decision-making uncertainty are not uncommon. Military conflicts and regulatory challenges imposed by new technological developments (e.g. cybersecurity) offer recurring examples of high-uncertainty policymaking situations. For the purposes of this paper, however, we focus on financial policy discussions in Congressional hearings before, during, and after the 2007-2009 Financial Crisis. Compared with other externally-imposed policymaking challenges, the Financial Crisis offers the largest and most recent example of a crisis event in the United States, and provides a useful opportunity to study whether this sort of break-down in ideological understandings of politics occurred.[5]

In order to capture the pre-crisis state of legislative decision-making, we examined all Congressional hearings on the issues of financial regulation, economic policy, and housing policy (as housing lending practices were the proximate cause of the collapse). To study changes in the structure of policy arguments and information, we employ a text-as-data approach that allows us to assess virtually all arguments and evidence set forth in formal committee hearings covering topics relevant to the crisis. Hearings are not the only mecha-

[4] Jones and Baumgartner [2005] term this disruption "attribute intrusion."

[5] Proof of a break-down in ideological understandings of the financial crisis occurred in at least example: neither the first nor the second vote on TARP could be understood via the Poole-Rosenthal scaling system for congressional votes [McCarty et al., 2013].

nism by which Congress processes information, but they are the place where the issues and arguments are put on the record. During hearings, the findings and recommendations of federal agencies, congressional analytical bureaucracies, committee staff, as well as the testimonies of interest groups and independent experts and the commentary of legislators are all recorded and made publicly available as text data, making them accessible for our approach. Usefully, the focus on hearings also allows us to employ the Policy Agendas Project's categorization system to structure our search for relevant documents, allowing us to construct an appropriate corpus relatively easily.

3 Dataset and Measurement

3.1 The Financial Crisis

In the years preceding the Great Recession, the financial sector underwent a quiet but steady revolution. Following the piecemeal repeal of Glass-Stegall, investment and commercial banks were allowed to merge to a higher degree than previously allowed. These institutions made substantial profits by exchanging high-risk securities and assets pooled together into complex financial products, whose risks were often not fully understood by financiers, nor disclosed to investors. One of the consequences of this relatively unconstrained regulatory environment was that the financial sector, as a whole, was susceptible to systemic risks if any single institution failed.

This looming risk went unknown or unacknowledged by most members of Congress. In one hearing of the House Financial Services Committee, Spencer Bachus declared that the "the economy appears strong and vibrant, and absent some unforeseen shock, likely to remain so."[6] Starting in the mid-to-late 2000s, however, a series of systemic shocks made the extent of these vulnerabilities clear. In November of 2007, destabilization in the subprime mortgage

[6] Spencer Bachus, U.S. Congress, House, Financial Services Committee *Monetary Policy and the State of the Economy, Part I*, 110th Cong., 1st Sess., February 15, 2007.

sector led lenders like Wells Fargo and Fannie Mae to write off billions, leading to soaring mortgage payments and foreclosure rates [Dennis, 2007, Ellison, 2007].

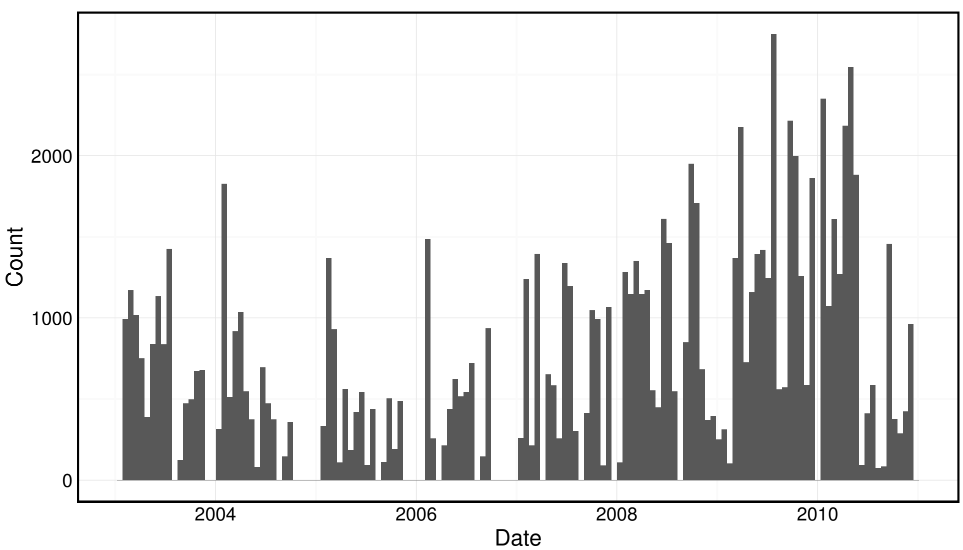
The banking sector faced a similar upheaval shortly afterwards. Pressed by the Federal Reserve Bank of New York, JP Morgan Chase acquired Bear Stearns to prevent its collapse in March 2008. The Federal Reserve Board announced it would monitor the finance industry closely and provide liquidity as necessary [Federal Reserve Bank of St Louis, 2015]. On September 7th, 2008, the federal government bailed out Fannie Mae and Freddie Mac, who were together responsible for almost half of all outstanding home mortgages. A week later, on September 15th, Lehman Brothers Holding Inc., a global financial services firm, filed for Chapter 11 bankruptcy, setting off a cascade of failures, bankruptcies and acquisitions that reshaped the banking industry. Within the month, Bank of America bought out Merrill Lynch and the Federal Reserve provided AIG with $85 billion in bailout funds. On September 29th, the market registered a 778 point fall in the Dow Jones Industrial Average, the largest 1-day drop in history [Financial Crisis Inquiry Commission, 2011].

The Great Recession triggered substantial legislative activity. The International Monetary Fund [2009], for example, called the crisis the worst global recession since World War II. In response to the severity of the crisis, Congress passed several acts: the Troubled Assets Relief Program (TARP), a $700 billion financial bailout plan, the American Recovery and Reinvestment Act of 2009, an $830 billion economic stimulus package, and the Dodd–Frank Wall Street Reform and Consumer Protection Act. Throughout this process, the United States Congress heard from panicked stakeholders eager to communicate the impact of the crisis on various sectors of the economy, including small business, mortgage availability, and middle-class financial stability. In the course of these hearings, participants proposed an array of prescriptions and policy solutions, many of which were not realized in the eventual legislative record. The transcripts of these hearings thus give us a rare window into the process of information collection and into how legislators cognitively process the attributes of a problem before, during and after a monumental crisis.

3.2 Building the Hearings Data

To construct our dataset, we relied on the hearings data published by the Government Publishing Office (GPO). First, we scraped all hearing texts and descriptive data available on the GPO's website as of January, 2015, including the names of witnesses and members of Congress, partisan affiliations, and committee and hearing titles. We then cross-referenced the GPO files with ProQuest Congressional and the Policy Agendas Project (PAP) datasets, and identified all available hearings that were categorized by PAP as either Macroeconomics ($n = 129$), Community Development and Housing ($n = 46$), or Banking, Finance, and Domestic Commerce ($n = 407$). This process left us with a total dataset of 582 hearings. As shown in Figure 1, the number of statement counts varies by date – some days reflect high levels of discussion, while other days involve few hearings or span Congressional recesses.

Figure 1: Statement counts over time.



Next, we parsed each individual hearing file by segmenting each hearing into a series of author-statement pairs using regular expressions.[7] We then matched authors to individual-level descriptors, such as party affiliation and seniority (drawn from Stewart and Woon [2011] and the GPO's own pages). In most cases, the GPO transcripts identify each author

[7] The GPO publishes its hearing transcripts as plain text files, without any embedded HTML structure or speaker-level information.

Figure 2: Statement counts by Policy Agendas Project major topic code.

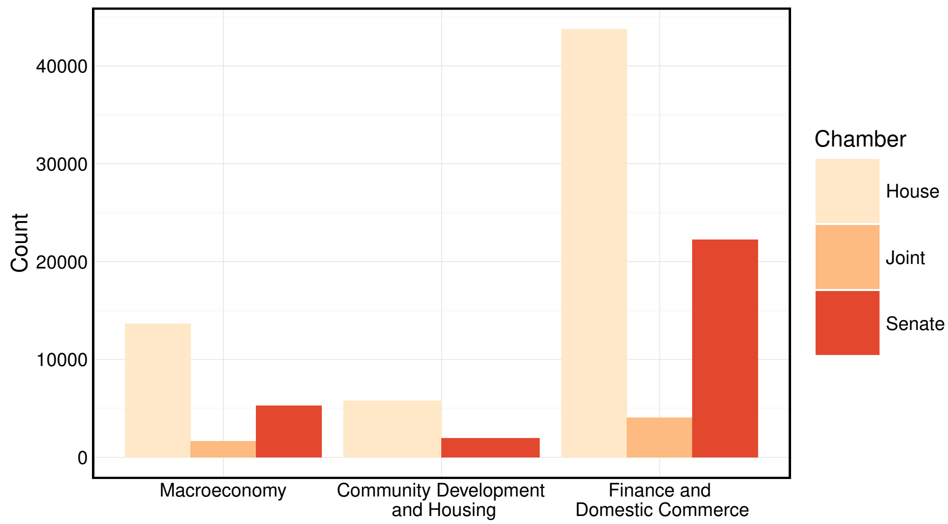
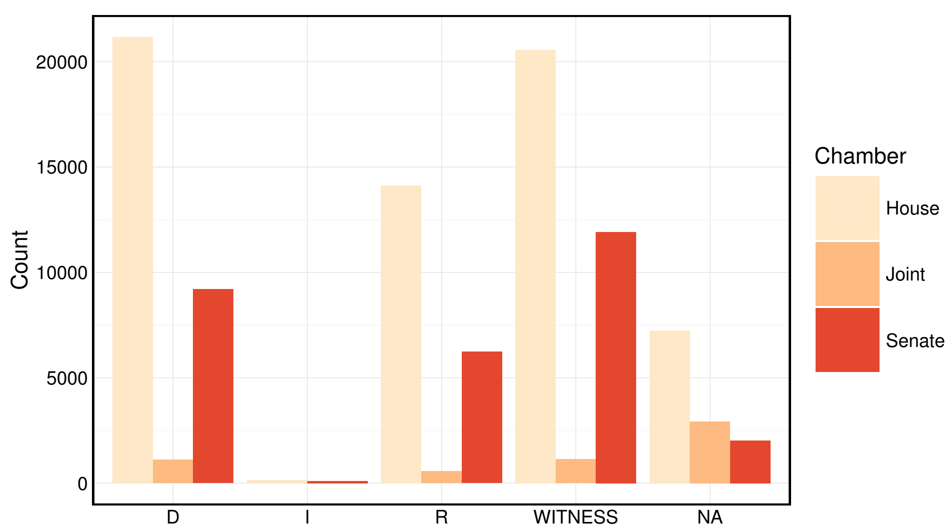


Figure 3: Statement counts by speaker type and chamber.



"NA" refers to speakers that were not matched to metadata entries.

using only the author's last name, which created substantial difficulties when attempting to uniquely identify individual speakers.[8] However, through a series of heuristics,[9] we were able to positively match 87% of statements in the dataset to at least some metadata. Figure 3 reports basic frequency data on these statements, including the number of statements by the type of speaker and by the chamber of the hearing. Wherever possible, we also discarded all prepared statements, written materials, and editor's notes from the transcript, leaving only the spoken contents of the text.

Finally, we conducted a series of standard preprocessing steps on the corpus.[10] In particular, we discarded punctuation, transformed all letters to their lower-case representations, and discarded stopwords[11] and words that were three characters or shorter. We also discarded words that appeared in fewer than 10 documents and documents with 5 or fewer words. This process left us with a final corpus of approximately 98,000 statements, 16,000 unique tokens, and 4.2 million total tokens.

3.3 Dimensionality Reduction

As Grimmer and Stewart [2013] note, textual data are inherently high-dimensional. Even in written form, speeches and conversation transcripts communicate a complex array of ideas, which are labor-intensive for human coders to analyze in a systematic fashion. Researchers working on these kinds of datasets are often interested in discovering and measuring a smaller set of latent themes, which capture the core ideas in the dataset of interest. In the text-as-

[8]Witnesses, in particular, were difficult to match accurately. Because the GPO relies on individual committees to supply metadata for each hearing, witness lists for individual hearings were frequently unavailable, making it impossible to positively identify witness statements through an automated procedure. We experimented with using the positively-identified statements to train a witness/member classifier, but were unable to obtain a sufficiently high degree of accuracy to proceed.

[9]See replication materials and Appendix C for details. Loosely, we first checked the hearing metadata for unique name matches in either the list of members present at the hearing, or the list of witnesses testifying at the hearing. If no unique matches were present, we then checked for similar matches in the transcript's table of contents (using regular expressions to separate members from witnesses). If neither of these options produced a unique match, we retained the statement and speaker name, but did not associate metadata with the document.

[10]See, e.g., Grimmer and Stewart [2013] for an overview of preprocessing practices.

[11]Using the stopword corpus available in Python's NLTK library (http://www.nltk.org/).

data setting, Latent Dirichlet Allocation (LDA) is likely the most prominent dimensionality reduction approach of this type (see Appendix A for details). [12] LDA is an unsupervised, generative modeling approach used widely in linguistics and computer science [Blei and Lafferty, 2009, Blei, 2012], which has been extended to include a variety of additional information sources and data structures (see, e.g., Blei and Lafferty [2006], Grimmer [2010], Lucas et al. [2015], Roberts et al. [2014], Teh et al. [2012]). As with other unsupervised approaches, LDA is designed to uncover some latent thematic structure underlying a collection of texts (e.g., newspaper articles or conversation transcripts). The latent variables generated by the model, which are usually known as *topics*, are discovered automatically based on patterns of word co-occurrence in documents, and then interpreted by subject experts in the relevant field. In political science, topic models and related modeling approaches have been applied to Senatorial press releases [Grimmer, 2010], Arabic fatwas [Lucas et al., 2015], and many other datasets.

In our case, we use the LDA topic outputs to represent the attribute space present in financial policy discussions (see Lam and Chan [2015] for a similar use of LDA). As with most data reduction-type models, LDA requires the user to supply the number of topics. We experimented with topic values between 30 and 100, but settled on 40 topics for the results presented in this paper, as inputs near this value seemed to offer the best balance between excessive subdivision of substantive themes and complete representation of topical content. For robustness, we repeated all analyses over other topic values; however, the choice of number of number of topics did not substantially affect our results.

As with any unsupervised modeling approach, constructing generalizable measurement schemes based on LDA topics is potentially challenging. In the supervised context, researchers can specify particular categories of interest, and train human or machine coders to identify sections of text associated with these themes. As usage shifts across time periods

[12] Using a Python wrapper available in the Gensim library [Řehůřek and Sojka, 2010] for the Mallet LDA implementation McCallum [2002]. We used default hyperparameter settings, including automatic optimization for the Dirichlet concentration parameter (updated every 20 iterations). With 4-fold parallelization and 10,000 iterations, this model took approximately an hour to train on a standard laptop computer.

and substantive settings, coders can adapt definitions to match new patterns and substantive concerns. By contrast, variables identified by unsupervised models are not guaranteed to match any particular conceptual scheme, and are more dependent on the dataset under consideration. As described in subsequent sections, we validate the topics we identify by examining a series of "exemplary" documents from selected topics, to ensure that the topics extracted contain policy-relevant language. However, we emphasize that the set of themes we present here are limited to the particular dataset we examine, and cannot be straightforwardly generalized to other settings.

3.4 Measuring Uncertainty

To quantify uncertainty in this domain, we first created a normalized word assignment vector for each speaker/hearing combination, consisting of the proportion of words uttered by a given speaker in a given hearing which were assigned to a given topic. Then, we calculated normalized information entropy values for each word assignment vector [following Boydstun et al., 2014]. Formally, normalized information entropy can be viewed as the expected information contained in an event or series of events, and is defined as:

$$\eta = \frac{1}{log(n)} \sum_{i=1}^{n} p_i log\Big(\frac{1}{p_i}\Big)$$

For $0 < p_i < \infty$, with n total bins (here, topics).

Intuitively, information entropy can be viewed as a measure of the "spikiness" (or concentration) of a finite-length vector of positive real numbers. For the purposes of this study, we use entropy to measure the extent to which a given person's statements in a particular hearing are devoted to a single topic. An entropy value of 0 implies that every word uttered by a particular person addressed the same topic, while a value of 1 implies that person distributed their words evenly between all topics.

As demonstrated in Boydstun et al. [2014], entropy is highly non-linear. In particular, small changes at the high end of the scale imply much larger distributional shifts than similar-sized changes at the low end of the scale. To aid interpretation, we therefore define the effective number of topics:

$$\tau = n^{\eta}$$

As shown by **?**, for any η, τ can be interpreted as number of evenly-distributed categories required to produce the given value of η (see Appendix B for proof and details). For example, for $n = 40$ topics, if a particular document divided its verbiage between those categories such that $\eta = 0.5$, a second person who divided their verbiage evenly between $\tau = 40^{0.5} \approx 6.32$ topics would also produce an equivalent entropy value $\eta = 0.5$.

In all subsequent analyses and modeling work, we operate on the transformed effective-topics statistic rather than the non-linear informational entropy scale. This transformation thus provides a more intuitive understanding of the entropy statistic, allowing readers to interpret changes in entropy on a linear scale and in terms of a more concrete construct.

3.5 Validation

As in most topic modeling applications, the topics produced in our case reveal a mix of substantively-interpretable and "junk" topics (see Figure 4 for top-probability words in selected topics). Generally, it is difficult to identify the dividing line between substantively interpretable topics and uninterpretable ones based on the top-probability words in a standard LDA output. In this case, our existing familiarity with the practical organization and rhythms of Congressional hearings guided us to classify topics as substantive if their top-probability words directly referred to public policy. Conversely, we identified topics as "junk" if they were comprised of an incoherent series of adverbs and common nouns (people, things, then, question, don't, etc.) or included language dealing predominantly with congressional procedure (gentleman, minutes, record, time, etc.).

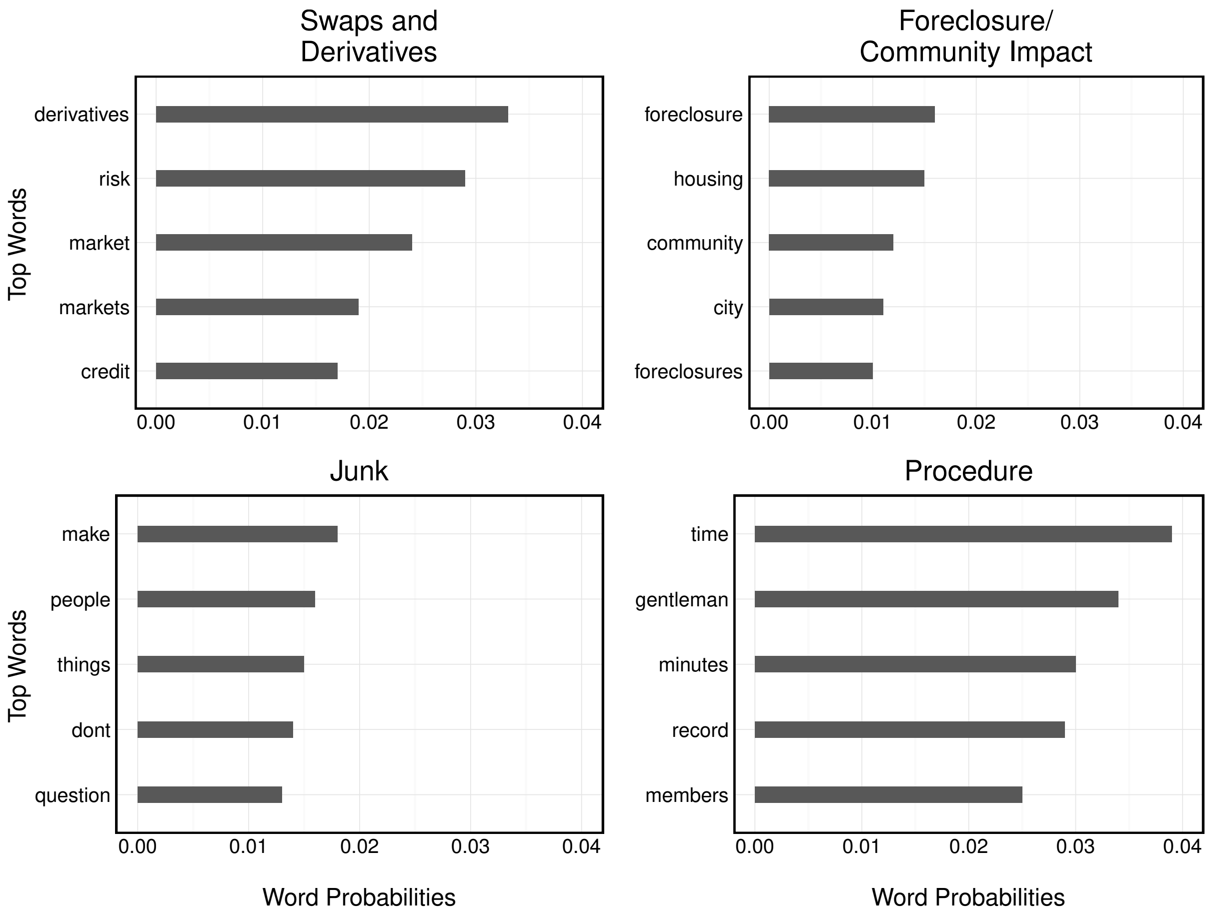


Figure 4: Top-probability words from selected topics.

Out of the 40 topics produced by the model, we identified 31 that were substantively important to our analysis. We assigned each topic a label according to the dominant policy addressed in their top-probability words. These included labels such as "Swaps and Derivatives," "Bank Regulation," and "Risk/Capital Requirements," among others. In our subsequent analyses, we experimented with both including and excluding "junk" topics; however, the choice between these options did not affect our results.

To further validate our topic labels and uncertainty measurements, we returned to the original text of a sample of hearings to validate that the topics identified by our model corresponded to the genuine substance of individuals' statements. We began by identifying statements with the lowest number of effective topics, indicating that the statement was focused on a single topic area. We then read the statement and those surrounding it to confirm that the statement corresponded to our chosen topic labels. Examples of these kinds of statements are given in Table 1. For each of the 31 substantively relevant topics we identified, we used these exemplar documents to adjust the labels we assigned based on top-probability words, and used these adjusted titles in our subsequent analyses (for a full list of all finalized topic labels refer to Figure 6).

Whereas statements assigned a low number of effective topics should resemble the narrow policy statements highlighted previously, statements characterized by a high number of effective topics should address many different dimensions of the financial crisis and exemplify the type of uncertainty we hypothesize should be highest at the height of the crisis. Here, too, the results pass a face validity test. For example, at a hearing of the Banking Housing and Urban Affairs Committee on March 4th, 2008, many different potential factors and actors were blamed for their roles in the financial crisis. Senator Chris Dodd urged regulatory agencies not to become too focused on any one factor or another:

> These are subject matters that very few people understand, including, I would say this respectfully, our colleagues here. Despite their good intentions to really understand the totality of all of this, not to have a stovepipe mentality about it, sort of looking at these things in sort of separate funnels, failing to recognize the interrelationships that occur here and how all of this is critically important to our economic success. But we count on you. That is where really this has to be. And Jack Reed's point here, the culture of how you approach your public responsibilities, your regulatory responsibilities, are critically important.[13]

Our experience reading individual statements suggested by our modeling and measurement approaches convinced us that our strategy provided a reasonable proxy for our concept of interest, allowing us to move forward with our analysis.

[13] Christopher Dodd, U.S. Congress, Senate, Banking, Housing, and Urban Affairs Committee, *State of the Banking Industry*, 110th Cong., 2nd Sess., March 4, 2008.

Table 1: Exemplar statements from chosen topics.

Topic Label	Statement
Swaps and Derivatives	The vast majority of **credit derivatives** take the form of the **credit default swap**, which is a contractual agreement to transfer the **default risk** of one or more reference entities from one party to the other. They are the fastest-growing part of the OTC **derivatives** business and the source of a great deal of innovation. **Credit derivatives** arose in response to two needs in the financial industry. The first was the need to hedge **credit risk**. Prior to the existence of **credit derivatives**, lenders had a limited number of ways to protect themselves if the financial condition of a borrower were to deteriorate. One was to take collateral and the other was by selling the loan, which normally requires the consent of the borrower.[a]
Systemic Risk	Let me just say a few things about AIG. AIG is a huge, complex, global insurance company attached to a very complicated investment bank hedge fund that was allowed to buildup without any adult supervision, with inadequate capital against the **risks** they were taking, posing putting your **government** in a terribly difficult position. Your **government** made the judgment back in the fall that there was no way that you could allow default to happen without catastrophic damage to the American people. That judgment, I am sure, was the right judgment at the time. Today we are in a situation where the world is dramatically worse. You are seeing pressures across broad parts of the **economy** in the **financial system**.[b]

Statements given are those with a high proportion of their (pre-processed) words drawn from the given topic. Words that are among the top ten highest-probability words for the given topic are bolded.

[a] Robert Pickel, U.S. Congress, Senate, Banking, Finance and Urban Affairs Committee, *Reducing Risks and Improving Oversight in the OTC Credit Derivatives Market*, 100th Cong., 2nd Sess., July 9, 2008.
[b] Timothy Geithner, U.S. Congress, House, Ways and Means Committee, *President's Fiscal Year 2010 Budget Overview*, 111th Cong., 1st Sess., March 3, 2009.

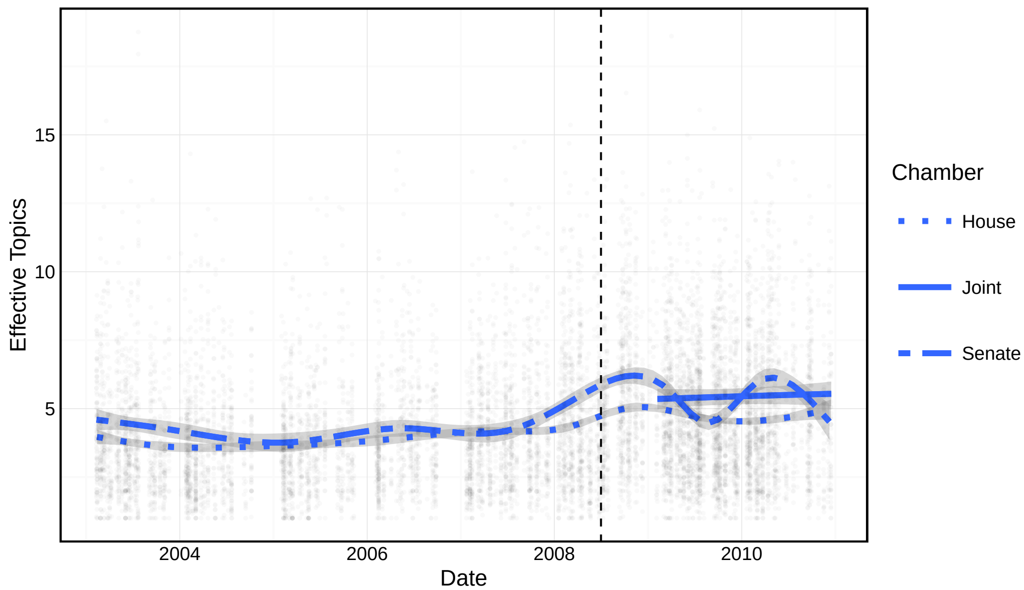
4 Results

4.1 Initial Findings

As hypothesized in the previous section, we find a substantial spike in effective topics corresponding to the 2008-2009 Financial Crisis. Unsurprisingly, baseline conversation diversity is relatively high throughout the period of study, which reflects the exploratory nature of most discourse in Congressional hearings. Despite this high baseline, however, we still detect a large increase in conversation diversity following the financial crisis. As shown in Figure 5, individuals participating in Congressional hearings during the post-crisis period (defined as the period after July 1, 2008, the date in which the US economy officially entered a recession) discussed approximately 0.9 extra effective topics on average compared to their pre- and post-crisis counterparts. This increase is not attributable to "junk" topics; rather, conducting the same analysis described above on the 31 topics that appeared to be directly policy-related produces an increase of the same magnitude. With "junk" topics retained, this spike corresponds to a 13.9% increase in effective topics (6.5 vs. 7.4); with "junk" topics excluded, the percentage increase rises to 21.8% (4.0 vs. 4.9).

In terms of the individual topics, most of the increased uncertainty observed during the crisis period came from spikes in solution-specific and relatively technical technical topics (see Figure 6). Of the topics with the largest crisis-period increases, most were either related to particular policy proposals (e.g. "TARP/Bailouts") or related to new problems and solutions that arose during the financial crisis period (e.g. "Swaps and Derivatives", "Foreclosure Prevention", "Systemic Risk"). This result is not surprising, and offers a useful face validity result for our analytical strategy. If members of Congress actually do face increased uncertainty during crisis periods, most of their attention should be focused on a (relatively diffuse) set of topics that capture the new policy issues created by the crisis. In the financial crisis dataset, we see exactly this pattern. Compared to the baseline period, members of Congress appear to be focusing on a larger, more diverse set of issues, which were mostly

Figure 5: Effective topics values by chamber.



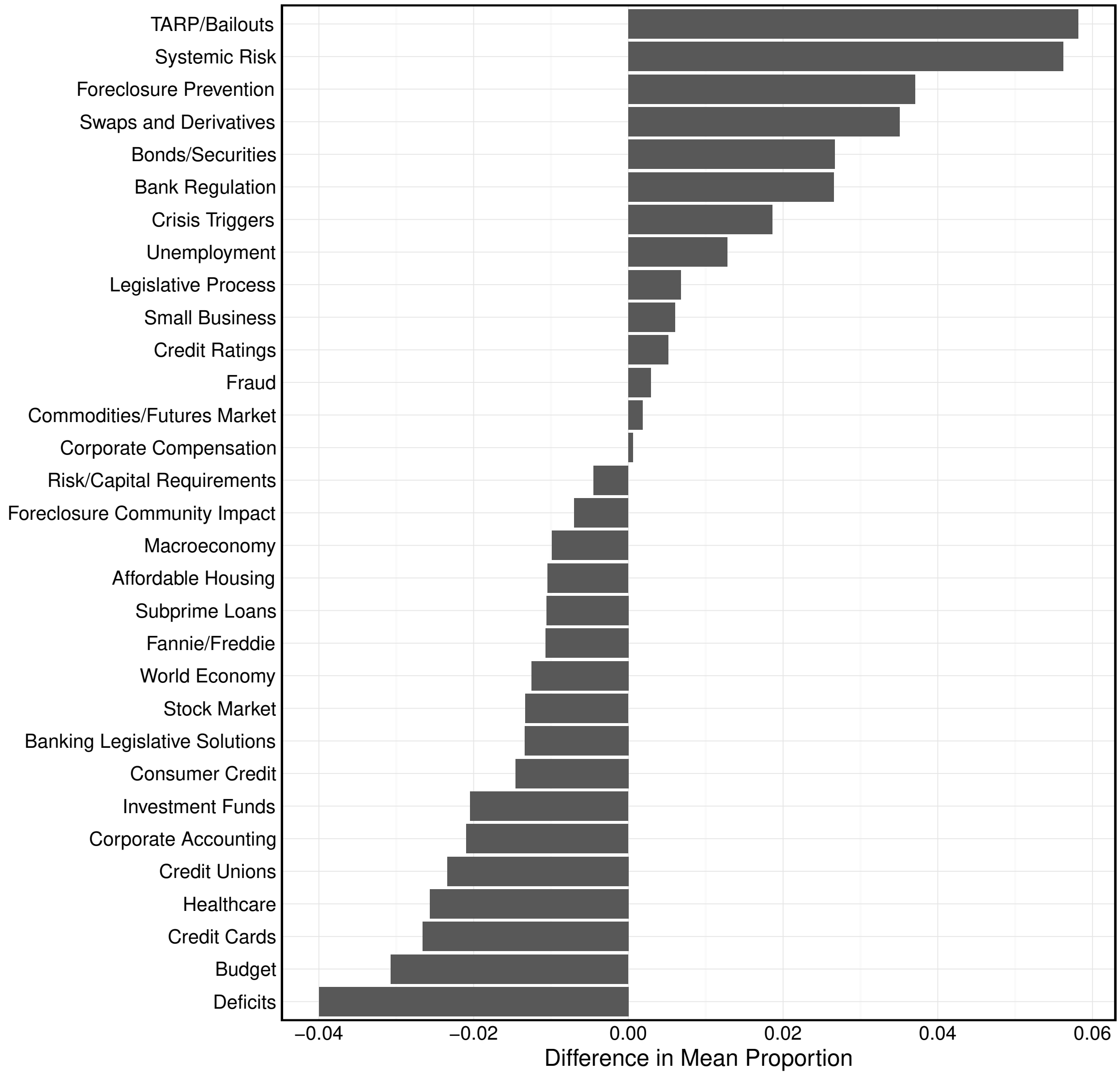
Bivariate smoother (via generalized additive model) fit to chamber-level subgroups, with 95% confidence intervals. Dashed line indicates July 1, 2008, the official start of the recession in the United States.

related to the new, relatively technical policy areas that were placed on the forefront of the agenda during the crisis period.

Generally speaking, the crisis-period effective topics spike appears to cut across most subgroups in the data, though with some important between-group differences. For example, speakers in Senate hearings exhibited a slightly larger crisis-period entropy spike than their House counterparts, with a much more noticeable peak just before the 2008 national election. In both chambers, committee chairs and ranking members in both chambers generally showed the highest overall entropy values in the dataset (see Figure 7). Compared with backbench members, committee chairs and ranking members are more conversant in their subject matter and more interested in appealing to a national constituency than many of their colleagues. As a result, we should expect these members to discuss a relatively broad range of topics.

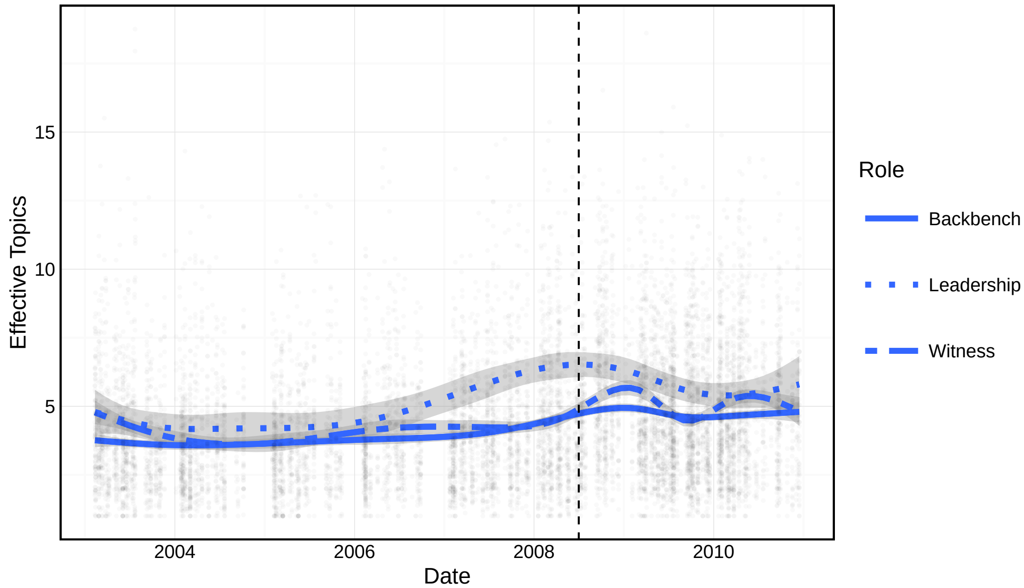
Witnesses occupy a similar position. Throughout the period we examine, witnesses discuss a consistently higher number of effective topics than backbench members, and are often nearly equal to members in leadership positions. In this case, discourse patterns within committee hearings likely provide the explanation. Outside of opening statements and other

Figure 6: Shifts in mean proportion of words by hearing for each topic, post-crisis.



introductory remarks, members of Congress spend most of their time during hearings asking short questions and making focused remarks. Witnesses, on the other hand, tend to address and expand on ideas raised by members, broadening the topical content of their statements in response to member prompting. During a crisis period, this prompting pattern is likely to be particularly acute, as members seek out broader types of information and force witnesses to address a wider range range of questions.

Figure 7: Effective topics values by speaker type.



Bivariate smoother (via generalized additive model) fit to chamber-level subgroups, with 95% confidence intervals. Dashed line indicates July 1, 2008, the official start of the recession in the United States.

5 Modeling

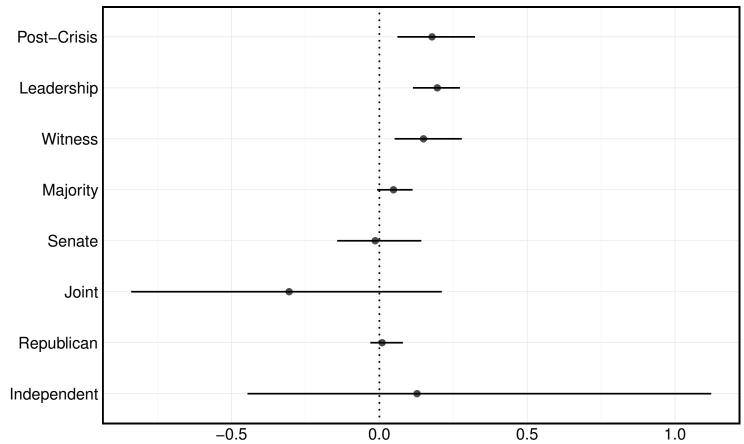
5.1 Results

Using the approach described above, we fit a two-level Bayesian beta regression model (see Appendix D for details and convergence assessment) to the effective topics data extracted using the 40-topic model ("junk" topics excluded, leaving 31 total topics), with a separate set of partially pooled coefficients estimated for each committee.[14] In addition to the pre/post Financial Crisis dummy variable (our main variable of interest), we also estimated coefficients for speaker type (Democrat, Republican, Independent, or Witness, using Democrat as the baseline), leadership status (coded as 1 if a speaker was a committee chair or ranking minority member and 0 otherwise), majority status (coded as 1 if the speaker was a member of the majority party in a given chamber in a given session of Congress) and the chamber in which the hearing was held (House, Senate, or Joint, using House as the baseline). As discussed

[14] As before, we also fit versions of this model using data drawn from LDA models estimated using 20, 40, 60, 80, and 100 topics, with "junk" topics included and excluded in each variant. However, we found that our results did not vary substantially across any of these model variants.

below, these additional variables were included to capture variation introduced by structural features of hearing organization, which offer certain speakers an opportunity to discuss a more diverse set of issue areas than others.

Figure 8: Top level coefficient posterior means.[a] (95% CI).

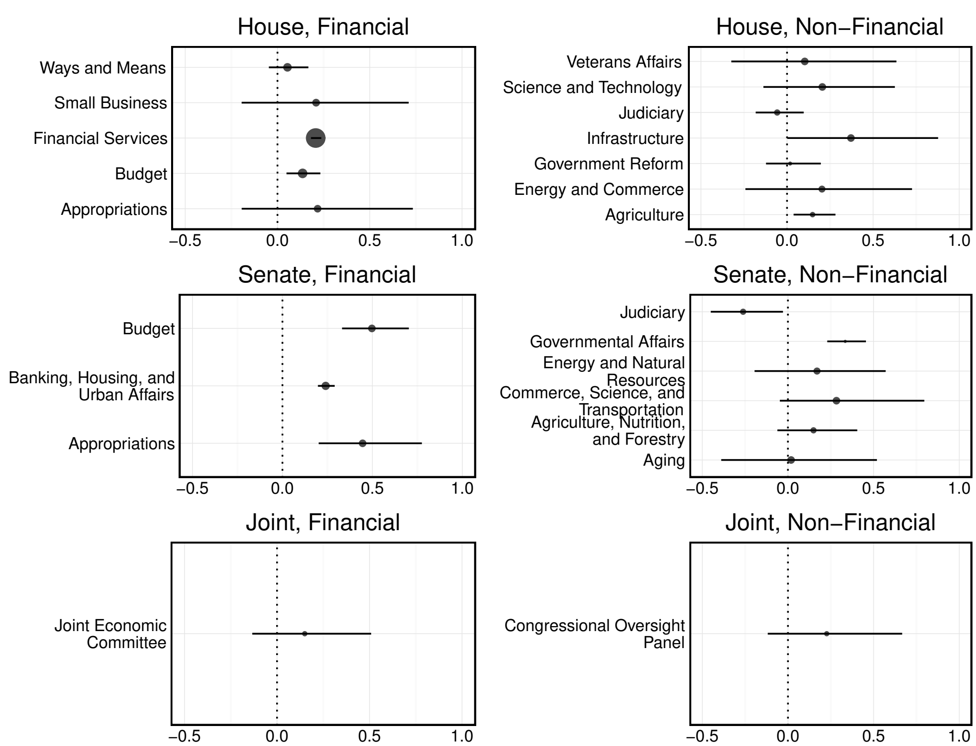


Lines indicate 95% credible intervals for given coefficients. "Democrat" used as baseline for "Republican" and "Independent," "Backbench" used as baseline for "Leadership" and "Witness," and "House" used as baseline for "Senate" and "Joint." "Post-crisis" is an indicator variable for statements issued after July 31st, 2008, the date at which the US economy officially entered a recession.

The results of this comparison are given in Figures 8, 9, and 10. Unsurprisingly, post-crisis statements and statements uttered by witnesses and speakers acting in a leadership role generally contain a substantially higher number of effective topics than their respective baselines (see Figure 8). As usual in the GLM setting (Bayesian or otherwise), effect sizes are not directly interpretable on the scale of the dependent variable. In this case, however, the posterior mean for the coefficient on post-crisis statements corresponds to roughly a 10% increase in effective topics for witnesses and a 20% increase in effective topics for statements made by leaders relative to the backbench baseline. Statements made after the financial crisis exhibit an increase of similar magnitude, corresponding to approximately a 20% increase in effective topics compared with those uttered before the start of the crisis.

The hierarchical modeling setup described earlier also allows us to explore effect sizes by committee. Encouragingly, for most variables in the model, coefficient sizes vary relatively little across the partially pooled committee-level coefficients. The leadership coefficient (shown in Figure 10) offers a representative example, with most other variables exhibiting a similarly low level of variation. Intuitively, the pattern seems reasonable. Besides the financial crisis variable (discussed below), most other variables with their posterior mass concentrated far from zero are related to structural features of Congressional hearings, which are unlikely to vary substantially by committee or policy area. For example, in most committee hearings, presiding members set the agenda for the hearing in a lengthy opening statement, which allows them to discuss a much broader range of topics than most other members. Witnesses, for their part, are often called upon to address a wider range of topics than individual members, and are also likely discuss a wider range of topics than backbench members. As a result, we should expect the coefficients on these variables to be consistently positive.

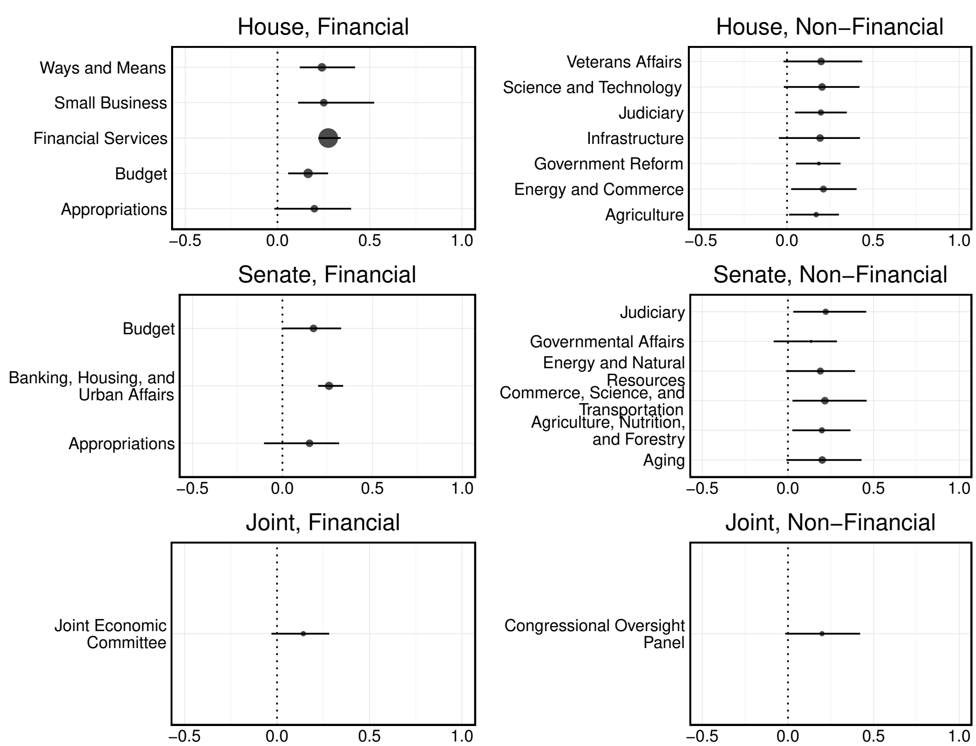
Figure 9: Committee-level leadership coefficient posterior means.



Lines indicate 95% credible intervals. Circle sizes are approximately scaled to the number of observations from each committee.

By contrast, the effect of the post-financial crisis coefficient is much more variable. Unsurprisingly, the coefficient posterior means are positive for most committees; however, only a few committees have their posterior density largely concentrated above zero. Encouragingly, these committees that exhibit the largest post-crisis coefficients are generally those committees that deal directly with financial policy (e.g. House Financial Services, Senate Banking and Financial Affairs, and House and Senate Budget). This pattern fits with the changes in topic usage documented previously. As noted earlier, the topics that received the most additional attention following the financial crisis were generally those that dealt directly with financial instruments and other technical matters (e.g. "Swaps and Derivatives" and "Systemic Risk"). Intuitively, committees dealing directly with financial policy are the most likely to hold hearings on these new issue areas, and to expand the number of issue areas under discussion accordingly.

Figure 10: Committee-level leadership coefficient posterior means.



Lines indicate 95% credible intervals. Circle sizes are approximately scaled to the number of observations from each committee.

6 Conclusion

In the US Congress, committee hearings serve a critical pre-decisional exploratory function. Compared with other information processing events, hearing transcripts offer a unique opportunity to examine patterns of discussion and information ingestion among lawmakers [Baumgartner and Jones, 2010]. In this paper, we have introduced a new approach to studying the processing of argumentation in public debates. We began with an essential thesis drawn from the existing literature: the debate conducted in hearings should expand in complexity during a crisis. As limitations in pre-crisis assumptions and framing patterns become apparent, policymakers are forced to expand their information ingestion efforts, broadening the set of possible problem and solution frames under consideration. As the crisis subsides, we hypothesize that the complexity of the debate space will decline as a "new normality" is established in the structure of argumentation.

To examine this thesis, we first developed a dataset consisting of all parsed, metadata-associated hearing transcripts made available through the Government Publishing Office's website, which we subsequently narrowed to a set of hearing transcripts identified as relevant to financial policy by the Policy Agendas Project. Earlier studies have treated each hearing as an indicator of the issues involved; here, we treat the actual statements made in the hearings as the indicators, allowing us to examine individual-level patterns. We then fit a topic model to this dataset, and examined patterns in growth and decline of conversation diversity following the 2007-2008 Financial Crisis (as measured by the effective topics statistic we introduce). We find a clear and substantively important increase before and during the crisis, with a subsequent decline as the crisis abated. However, this effect is concentrated in committees whose jurisdictions deal wholly or partially with financial policy. Non-financial committees, by contrast, exhibited relatively limited shifts in discourse patterns in response to the crisis. Finally, we also find important differences in discourse patterns among various types of participants. Leadership members, in particular, consistently address a more diverse array of topics than their backbench counterparts.

Besides its substantive contributions, this paper also provides a novel methodological approach to the study of discourse patterns in public debates. By focusing on conversation diversity rather than on patterns of change in particular themes, we are able to avoid many of the generalization challenges that plague standard topic modeling applications. Analyses of large textual datasets offer vast potential, but thus far relatively little attention has been directed at tracing complex processes through time. We offer a start on this enterprise.

References

Bethany Albertson, John Brehm, and R Michael Alvarez. Ambivalence as internal conflict. In *Ambivalence and the structure of political opinion*, pages 15–32. Springer, 2005.

R Michael Alvarez and John Brehm. American ambivalence towards abortion policy: Development of a heteroskedastic probit model of competing values. *American Journal of Political Science*, pages 1055–1082, 1995.

Frank R Baumgartner and Bryan D Jones. *Agendas and instability in American politics*. University of Chicago Press, 2010.

Frank R Baumgartner and Bryan D Jones. *The Politics of Information: Problem Definition and the course of public policy in America*. University of Chicago Press, 2015.

Frank R Baumgartner, Christian Breunig, Christoffer Green-Pedersen, Bryan D Jones, Peter B Mortensen, Michiel Nuytemans, and Stefaan Walgrave. Punctuated equilibrium in comparative perspective. *American Journal of Political Science*, 53(3):603–620, 2009.

David M Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.

David M Blei and John D Lafferty. Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pages 113–120. ACM, 2006.

David M Blei and John D Lafferty. Topic models. In Ashok Srivastava and Sahami Mehran, editors, *Text mining: classification, clustering, and applications*, pages 71–93. CRC Press, Boca Raton, FL, 2009.

Amber E Boydstun, Shaun Bevan, and Herschel F Thomas. The importance of attention diversity and how to measure it. *Policy Studies Journal*, 42(2):173–196, 2014.

John T Cacioppo and Gary G Berntson. Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological bulletin*, 115(3):401, 1994.

Bob Carpenter, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *J Stat Softw*, 2016.

Dennis Chong and James N Druckman. Framing theory. *Annu. Rev. Polit. Sci.*, 10:103–126, 2007.

Michael D Cohen, James G March, and Johan P Olsen. A garbage can model of organizational choice. *Administrative science quarterly*, pages 1–25, 1972.

Stephen Dennis. Democrats seek edge on mortgage issue. November 15th 2007.

Keith Ellison. What is the outlook for alleviating the mortgage crisis in the final year of the 110th congress? crisis rooted in economic insecurity. December 3rd 2007.

Federal Reserve Bank of St Louis. The financial crisis time line. `https://www.stlouisfed.org/financial-crisis/full-timeline`, September 14th 2015.

Silvia Ferrari and Francisco Cribari-Neto. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815, 2004.

Financial Crisis Inquiry Commission. The financial crisis inquiry report: Final report of the national commission on the causes of the financial and economic crisis in the united states. Technical report, Washington, DC, 2011.

Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992.

Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis.* Taylor & Francis, Boca Raton, FL, 3 edition, 2014.

Justin Grimmer. A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis*, 18(1):1–35, 2010.

Justin Grimmer and Brandon M Stewart. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3):267–297, 2013.

Melvin Hinich and Michael Munger. *Ideology and the theory of public choice*. University of Michigan Press, Ann Arbor, MI, 1994.

International Monetary Fund. World economic outlook - april 2009: Crisis and recovery. Technical report, Washington, DC, 2009.

Bryan D Jones. *Reconceiving Decision-Making in Democratic Politics: Attention, Choice, and Public Policy*. University of Chicago Press, 1994.

Bryan D Jones and Frank R Baumgartner. *The politics of attention: How government prioritizes problems*. University of Chicago Press, Chicago, IL, 2005.

John W Kingdon. *Agendas, alternatives, and public policies*, volume 45. Little, Brown, Boston, MA, 1984.

Keith Krehbiel. Paradoxes of parties in congress. *Legislative Studies Quarterly*, pages 31–64, 1999.

Wai Fung Lam and Kwan Nok Chan. Policy advocacy in transitioning regimes: Comparative lessons from the case of harbour protection in hong kong. *Journal of Comparative Policy Analysis: Research and Practice*, pages 1–18, 2015.

Daniel Lewandowski, Dorota Kurowicka, and Harry Joe. Generating random correlation matrices based on vines and extended onion method. *Journal of multivariate analysis*, 100(9):1989–2001, 2009.

Will Lowe and Kenneth Benoit. Validating estimates of latent traits from textual data using human judgment as a benchmark. *Political Analysis*, page mpt002, 2013.

Christopher Lucas, Richard A Nielsen, Margaret E Roberts, Brandon M Stewart, Alex Storer, and Dustin Tingley. Computer-assisted text analysis for comparative politics. *Political Analysis*, page mpu019, 2015.

Andrew Kachites McCallum. Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu, 2002.

Nolan McCarty, Keith T Poole, and Howard Rosenthal. *Political Bubbles: Financial Crises and the Failure of American Democracy*. Princeton University Press, Princeton, NJ, 2013.

Allen Newell and Herbert A Simon. *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, NJ, 1972.

Keith T Poole and Howard Rosenthal. *Congress: A political-economic history of roll call voting*. Oxford University Press, New York, NY, 2000.

Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. `http://is.muni.cz/publication/884893/en`.

Margaret E Roberts, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand. Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4): 1064–1082, 2014.

Charles III Stewart and Jonathan Woon. Congressional committee assignments, 103rd to 112th congresses. 2011.

Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical dirichlet processes. *Journal of the american statistical association*, 2012.

Supporting Information

A Modeling Details

To investigate uncertainty patterns in the hearing transcript data more closely, we fit a two-level Bayesian hierarchical beta regression to the effective topics dataset described above. This modeling approach, which is largely drawn from standard frequentist formulations of the beta regression [Ferrari and Cribari-Neto, 2004], proceeds as follows:

$$\begin{aligned} \tau_{ij} &\sim Beta(\mu_{ij}, \phi_{ij}) \\ \mu_{ij} &= f^{-1}(X_{ij}^T \beta_j) \end{aligned}$$

Where X_i is a matrix of predictor variables and τ_{ij} is the effective topics value for the i^{th} speaker/hearing combination, nested in the j^{th} committee. Here, the Beta distribution is parameterized as in Ferrari and Cribari-Neto [2004]'s treatment, such that if $\tau \sim Beta(\mu, \phi)$ then $E(\tau) = \mu$ and $Var(\tau) = \frac{\mu(1-\mu)}{1+\phi}$, and f^{-1} denotes the inverse logistic link function. Parameterized in this fashion, the model is inherently heteroskedastic; usefully, though, this formulation also allows for direct modeling of the precision term if desired.

As usual in the Bayesian hierarchical modeling framework, we then place shared priors on the regression coefficients:

$$\beta_j \sim MVNorm(\gamma, \Sigma)$$

Where β is a JxK matrix of first-level regression coefficients, with J equal to the number of committees and K equal to the number of parameters in the individual-level mean model. Similarly, γ is a $Kx1$ matrix of second-level regression coefficients. We place priors on the variance-covariance matrix $\Sigma = diag(\nu)\Omega diag(\nu)$ using the strategy suggested by Gelman et al. [2014]:

$$\nu \sim Cauchy(0, 2.5)$$
$$\Sigma \sim LKJ(1)$$

With ν restricted to the non-negative reals, and with Cauchy parameters selected to produce a vague but mildly informative distribution for regression parameter variances. LKJ refers to the Lewandowski et al. [2009] correlation matrix distribution, which reduces to an identity distribution over covariance matrices with shape parameter equal to 1.

Finally, we place independent priors on the second-level regression coefficients:

$$\gamma_k \sim Normal(0, 10)$$

With parameters selected here to give vague but mildly informative distributions for regression coefficients on the logistic scale.

Note that this approach implies an important simplifying assumption; namely, that the variance-covariance matrix between the first-level regression coefficients are equivalent across all first-level groups (committees). This assumption eases computational challenges substantially and seems plausible in the context of Congressional hearings, which are organized similarly and follow similar discussion dynamics across all policy areas.

We ran this model for 20,000 iterations, and assessed convergence by inspecting $\hat{R}$ [potential scale reduction factor; Gelman and Rubin, 1992] and effective sample size statistics as given in RStan [Carpenter et al., 2016]. For nearly all identified parameters, $\hat{R}$ values were ≤ 1.05 and all effective sample sizes were ≥ 1000, with most values achieving substantially higher effective sample sizes and lower $\hat{R}$ values. In a small number of cases, committees with very small observation counts produced coefficient values with convergence diagnostics outside these ranges. Based on these results, we were sufficiently satisfied that convergence had been achieved to move forward with our analysis.