Power in Text: Extracting Institutional Relationships from Legal Language

Robert Shaffer
University of Pennsylvania
rbshaffer@utexas.edu

September 28, 2018

Abstract

How do legislators allocate policymaking authority? Formal allocation of power is often framed as a *political* problem. Under unified government, legislators should favor simple, efficient implementing arrangements, while under divided government they should favor complex systems that optimize oversight. By contrast, in this paper I argue that institutional design choices should be primarily shaped by the *policy issues* under consideration, with bills addressing more complex and publicly salient issue areas yielding correspondingly complex institutional solutions. Using a novel measurement technique - which leverages both modern machine learning methods and case-specific knowledge regarding Congress's internal drafting standards - I collect data on institutional design choices in all American legislation enacted from 1990-2014. I find that issue salience is positively related to institutional complexity on average, but that the magnitude of this effect is conditional on policy area. I also find that ideological differences between the executive and the legislature are at best weakly related to institutional design choices, which challenges results presented in previous studies.

1 Introduction

How do legislators allocate policymaking authority? As any would-be lawyer knows, statutes, constitutions, and other formal legal texts establish relationships between actors, describing who can do what, when, and to whom. However, the nature and complexity of the institutional structures created by these texts varies enormously. For example, in American government, the Clean Air Act grants the Environmental Protection Agency sole decision-making authority over most policy decisions, while the Patient Protection and Affordable Care Act (ACA) fragments decision-making authority between the Departments of Labor, Treasury, Health and Human Services, and Veterans' Affairs, as well as various subordinate agencies and other institutions.

Political scientists and legal scholars have generally framed the allocation of authority decisions contained in these statutes in *political* terms. When the executive's interests are aligned with those of the legislature (and are likely to remain so in the future), legislators are more willing to pass simple "framework"-style legislation involving one or a few implementing actors and simple decision-making rules. By contrast, when legislative and executive interests diverge (or might do so in the future), legislators tend to favor complex institutional structures with many implementing actors, which provide greater oversight opportunities by fragmenting decision-making.

Beyond these basic political factors, we should also expect formal institutional design choices to be affected by the issues and policy areas under consideration. For example, because designing complex institutional arrangements is a costly process, we should expect lawmakers to be more willing to pay these costs when addressing issues that are salient to their interest group and electoral constituencies. Similarly, we should expect lawmakers favor complex decision-making structures when addressing policy areas which are themselves complex, since these these problems demand cross-cutting institutional solutions involving multiple administrative agencies. Unfortunately, while issue and policy area variables have been discussed in the theoretical literature on formal allocation of authority, ideas like these are largely absent from empirically-oriented scholarship. Largely, this limitation results from a measurement problem. Even for motivated and well-resourced academics, reading and interpreting legal texts is labor-intensive. As a result, most empirical work on allocation of authority has been restricted to single policy areas or to small sets of "significant" legislation, restricting the theoretical scope of these projects.

In this paper, I address these limitations in two parts. First, from a conceptualization and measurement standpoint, I argue that formal legal texts are best viewed as *relational* documents, which describe which actors can do what, when, and to whom. I then implement a neural network-based approach designed to extract these "implementing networks", which I validate both qualitatively and in an out-of-sample quantitative test. Second, I use this measurement approach to study fragmentation of authority in an original dataset consisting of all enacted American laws passed from 1990-2014. As predicted, I find that legislators tend to create more complex institutional arrangements when crafting high-salience laws in policy areas that cut across multiple policy domains (e.g. macroeconomics) than when addressing lower-profile bills in more focused policy domains (e.g. public lands). More surprisingly, after controlling for policy and issue area I find that executive/legislative preference disagreements are largely unrelated to complexity of formal institutions. These findings challenge the existing literature on formal allocation of authority, and suggest exciting directions for future research.

2 A Relational Conception of Formal Power

2.1 Preferences or Policy?

When a policymaker writes a formal legal document - such as a constitution, law, or administrative regulation - that policymaker must choose between a range of possible institutional models with which to implement the policy programs outlined in that document. At one extreme, she could create a complicated, fragmented implementing structure, in which each decision requires joint approval from many actors. At the other, she could write a simple "framework" document, which delegates most authority to one or a few implementing actors. The choice between these two models is consequential; by dividing policymaking authority between several actors, lawmakers reduce implementing efficiency and limit implementer discretion, but offer more opportunities for outside groups to monitor and intervene into the policymaking process. Streamlined implementing structures, by contrast, offer a more efficient policymaking process but fewer oversight opportunities.

To explain variation in institutional complexity, most empirical work on formal institutional design focuses on *preference-based* factors. As Moe (1990a; 1990b; 2012) and Moe and Caldwell (1994) argue, complex implementing structures are not *ex ante* desirable. From a policy standpoint, complicated power-sharing arrangements curtail administrative flexibility, reducing implementer responsiveness and promoting policy gridlock. Worse, complex implementing structures are costly to create. If a lawmaker wants to design a complicated institutional structure involving many implementing actors, he or she needs to devote a greater quantity of time and attention than a correspondingly simple institutional structure would demand. However, if the policymaker and the implementer possess different substantive preferences - for example, during periods of divided government - complex institutional structures become more attractive. By creating overlapping decision-making processes that involve a larger number of actors, lawmakers can slow the implementation of undesirable policy programs and increase opportunities for "fire alarm"-style oversight (McCubbins and Schwartz 1984). Various authors have found empirical support for this proposition across a variety of contexts, ranging from national-level American legislation (Epstein and O'Halloran 1999; Farhang and Yaver 2016) to the European Community (Franchino 2004, 2007).

While persuasive, these studies also contain important gaps. In the formal modeling literature, a variety of authors (e.g. de Figueiredo Jr 2002; Bendor and Meirowitz 2004) argue that complexity should be *more* common under unified compared with divided government, since lawmakers enjoying temporary control over both the executive and legislative branches have an incentive to "insulate" their preferred programs through directive language and complex institutional structures. Huber et al. (2001) and Huber and Shipan (2002) find conditional support for this explanation in the context of state-level health policy. In states with sufficiently professionalized legislatures, these authors report a negative relationship between unified government and implementer discretion, as opposed to the positive relationship reported by the studies I cite above. Results like these suggest that the relationship between preference disagreements and downstream allocation of authority choices is at least more complicated than most existing studies suggest.

Perhaps most importantly, we should also expect formal institutional design to be influenced by characteristics of the *issues* and *policy areas* under consideration. Because institutional design is a costly process, we should expect lawmakers to be more willing to create complicated institutional structures when addressing issues that are salient for important constituencies. For the purposes of this paper, by "salient issues" I refer to those issues that capture publicly-expressed attention from lawmakers' constituents and civil society and partisan stakeholders. Issues can capture public attention for a variety of reasons; for example, in the United States, periodic negotiations like those over the Farm Bill create predictable spikes in public attention to agricultural policy, while exogenous shocks like the 2008 Financial Crisis create more erratic spikes in public attention to banking and macroeconomic concerns. But, whatever the source, increased public attention to a particular policy problem should prompt politicians to examine that problem more closely, and should render lawmakers more willing to craft complex implementing structures.

Formal institutional design choices should also be affected by the characteristics of the broad policy area under consideration in a particular law. In an abstract sense, some broad policy areas simply involve more complex policy considerations than other. For example, macroeconomics legislation often requires collaboration between actors with expertise in energy, mortgage policy, and agricultural futures, in addition to standard financial expertise. By contrast, bills distributing oversight authority over public lands often involve a narrower set of considerations, which can be administered by one or a few administrative agencies. Intuitively, we should expect formal institutional design to reflect the concerns of the underlying policy area, with laws addressing more complex policy problems containing more complex institutional structures.

Policy and issue effects like these are also likely to affect the relationship between standard political variables and downstream allocation of authority choices. Again, because institutional design is costly, we should expect lawmakers to be differentially motivated by preference disagreement when addressing high-salience policy problems compared with their lower-salience counterparts. When addressing low-salience issues, politicians have little incentive to carefully design implementing structures that implement their preferred policy outcomes. In these cases, the effects of executive-legislative preference disagreements on downstream institutional design choices should be muted. By contrast, when addressing high-salience policy problems the relationship between preference disagreements and downstream institutional design choices should be sharper, since politicians should be more willing to create formal institutional structures that implement their policy preferences.

Though intuitive, policy effects like these have not been investigated extensively in the existing empirical literature. The reason for this limitations is simple: *measuring text-based distribution of authority is difficult.* Even for experts, parsing legal texts is difficult and labor-intensive, forcing scholars to limit the scope of their studies. To underscore the extent of this measurement challenge, consider Farhang and Yaver (2016)'s recent study. In their paper, the authors read and coded some 24,000 pages of legislative text in order to produce data on some 366 laws passed from 1947 to 2008. Clearly, though the hand-coding methodology Farhang and Yaver (2016) employ in their project produced important insights, expanding the scope of their dataset is likely impractical.

These data collection difficulties have forced scholars to limit the scope of their work, in at least two important ways. Most prominently, few studies compared patterns in allocation of authority across multiple policy areas, preventing scholars from analyzing the relationship between policy complexity and downstream allocation of authority (see, e.g. Huber and Shipan 2002, which focus exclusively on healthcare policy). The few exceptions to this rule (e.g. Epstein and O'Halloran 1999; Farhang and Yaver 2016) are limited in an equally important fashion. In particular, these studies only include prominent legislative proposals (e.g. Mayhew (1991)'s list of historically "significant" legislation), without examining their lower-profile counterparts. In order to test for the kinds of policy effects I describe in this section, then, we need a more scalable measurement approach.

2.2 From Text to Networks

In the social sciences more broadly, a common way to study power relationships in a document of interest is to study the *relational* statements that document contains. For example, in political science, a notable example of this kind of approach is GDELT (Leetaru and Schrodt 2013), which mines news accounts for subject/action/object triples corresponding to international events. Franzosi et al. (2012) use a similar approach to study narratives of victimhood and agency in newspaper accounts of lynching episodes in the American South. Among other findings, this analysis reveals a surprisingly strained relationship between law enforcement officials and white lynching mobs, who targeted sheriffs and deputy sheriffs for violence in nearly half of all lynching episodes.

This approach is similarly helpful in the legal setting. Among other functions, legal texts articulate power relationships, describing who can do what, when, and to whom. As a result, a reasonable way to conceptualize a law, contract, or constitution is to view it as a *network* of institutional relationships. Nodes, in this view, represent actors involved in the execution of powers outlined in the text, while edges represent the relationships between them. Depending on the scope of the document, these relationships might be directed or undirected, and might include simple connection types or more complex ones. To take an example familiar to most Americans, the US Constitution provides the House and the Senate joint authority over "legislative Powers herein granted"[1] (an undirected relationship) while empowering the Senate alone to offer "Advice and Consent"[2] to the President over treaties, and judicial/executive appointments (a directed relationship). By collecting and combining all such entities and relational ties, we can construct an "implementing network" for the US Constitution, which describes the set of formal institutional relationships envisioned by the document.

Network representations of this kind also offer a straightforward way to define key concepts like institutional complexity. In their study, Farhang and Yaver (2016) define "fragmentation" - an idea closely related to complexity - as:

> Division of implementation authority over a larger number of distinct actors, over a larger number of different agencies, and giving multiple actors the authority to perform the same function with respect to the same statutory provisions.

From this perspective, an implementing structure is complex to the extent that it grants a larger number of actors overlapping responsibility over execution of a particular policy program. The more actors a particular text involves in a particular policy program - and the more that those actors are coinvolved in the individual components of that policy program - the more complex and fragmented that text becomes. As I show in the following section, both of these ideas can be naturally operationalized using standard network-analytic tools, offering additional value to the conceptual approach I propose.

3 Constructing Implementing Networks

Besides its conceptual advantages, a relational conceptualization of formal power also helps to pinpoint the measurement challenges involved in studies of this kind. At their most fundamental level, networks are constructed from *entities* (the actors involved in the network) and *edges* (the ties between them). Treating formal legislative texts as implementing networks implies two natural measurement problems: in particular, what is the set of

[1] U.S. Const. Art I, §1.
[2] U.S. Const. Art II, §2.

implementing actors named in a particular legal document, and what are the relationships between them? As I show in the following section, both of these quantities can be extracted in straightforward, scalable fashion using computational linguistics methods, which helps to address the scaling problems I raise previously.

3.1 Data

To investigate patterns of complexity in American legal texts, I constructed a dataset consisting of all legislation (and accompanying metadata) passed by Congress and available through congress.gov, the official U.S. Congress legislative database.[3] Temporally, this dataset approximately covers the period from 1990-2014.[4] I then merged this dataset with the Congressional Bills Project's metadata set, which I used to filter "unimportant" bills from the dataset. As defined by the Congressional Bills Project, "unimportant" laws include "commemorative" laws and laws which transfer small quantities of land between government entities, and are therefore unlikely to include allocation-of-authority language. After removing these bills, this process left me with a final dataset of 3467 observations (out of 4800 total).[5]

Before proceeding, I conducted some simple preprocessing steps on each document. For each text, I first stripped any administrative headers (e.g. date of passage; legislative history; transcription notes), and segmented each document into sections.[6] I then removed the first section from each document. In contemporary American legislation, the first section of each law always contains a set of preliminary material - such as an official "short title" or a table of contents - which is not relevant for the analytical task I undertake in this paper.

[3] The tools used to construct this database are available via Github as Legislative_Data.

[4] At the earlier part of this period, congress.gov's coverage is not complete. As a result, conclusions drawn from this period should be interpreted with caution.

[5] Additionally, 42 observations from other parts of the dataset were missing due to metadata errors in the Congressional Bills Project's data.

[6] Using the regular expression parser I implemented in constitute_tools, a set of utilities I designed to assist with Constitute's data collection efforts. This parser separates each document according to a given set of organizational headers (e.g. titles; sections), while maintaining the internal hierarchy of each document. See Appendix A for details and sample parsed text.

3.2 Entity Extraction

Entity extraction is a classic natural language processing (NLP) problem, which has been attacked using a variety of heuristic and machine-learning approaches. In political science, perhaps the most common approach is a dictionary-based system (see, e.g. Leetaru and Schrodt 2013), in which the entities of interest are pre-identified using a dictionary generated by expert researchers. This approach generally produces few false negatives, but misses a large number of items of interest, since generating a comprehensive named entity dictionary is impractical in most situations. By contrast, general-purpose machine learning approaches - which rely on lexical and grammatical information to make tagging decisions - usually capture most entities of interest, but will also include irrelevant entries such as the names of people or places (Manning et al. 2014). For the purposes of this paper, I therefore opt for a localized machine learning approach, which splits the difference between these two alternatives by leveraging expert-generated lists of government entities to train a specialized named entity recognition model. This approach consists of three steps - *dictionary assembly*, *training set construction*, and *model fitting* - which I describe in detail in this section.

To train the localized model I use in this paper, I first built a custom dictionary of institution names that are likely to be present in American legislative texts. I began by scraping all names contained in usa.gov, the Federal Register, or one of five Wikipedia sources: specifically, the lists of federal agencies, defunct federal agencies, House committees/subcommittees, Senate committees/subcommittees, and joint committees. I then removed common prefixes and suffixes from these items (e.g. "United States"; "USA"), and stripped names of states and national governments (e.g. "Texas"; "California"; "Federated States of Micronesia") from the list. As an additional quality control measure, an undergraduate research assistant read a random sample of legislative texts, and supplemented this list with a series of additional missing items. The final dictionary produced by this process contained some 1360 items, representing most prominent institutions contained in the executive and legislative branches in American government.

Next, I used this dictionary to identify in-context examples of named entity mentions in actual American legislation. Using the American legislation dataset I describe in §3.1, I first segmented the text into sentences.[7] Since the computational complexity of the LSTM model I use scales with the length of the longest input sentence, to ease computation I then discarded all sentences longer than 75 words, leaving me with a training set consisting of some 29,080 sentences. For each sentence remaining sentence, I then conducted a simple string search for each named entity contained in my entity dictionary. If a particular entity was present in a particular sentence, I marked the first token of the entity with a "B-MISC" tag (denoting the beginning of the named entity), and any subsequent tokens with an "I-MISC" tag (denoting words inside the named entity). Finally, I marked all tokens not identified using one of these labels with an "O" tag.[8] Recovering these tags - and applying analogous tags to out-of-sample training examples - therefore represents the objective for the machine learning model I employ in this section.

To reach actual tagging decisions, I use a long short-term memory (LSTM) neural network, which is a common approach in modern NLP work.[9] Broadly, neural networks are a class of machine learning approaches which seek to predict some outcome of interest based on a series of interconnected "hidden layers". Each "hidden layer" consists of a set of unobserved parameters, which are generated as a weighted sum of the previous layer's parameters filtered through a non-linear "activation function" (such as a sigmoid or hyperbolic tangent function). These weights are iteratively optimized through stochastic gradient descent or a related algorithm to minimize prediction error. Recurrent neural networks - of which LSTMs are a variant - build on this framework by allowing the predicted output for a given data point to be influenced by the predictor variables as well as the corresponding predicted outputs for "adjacent" data points. This approach creates a recursive, context-sensitive prediction

[7] Using the pretrained Punkt sentence tokenizer, available via NLTK.

[8] Since some named entities are substrings of others - for example, compare "Secretary of Defense" with "Assistant Secretary of Defense" - before searching each sentence I ordered the named entity dictionary from longest tag to shortest, to ensure that the longest present named entity would be tagged first.

[9] E.g. language modeling (Sundermeyer et al. 2012) and part of speech tagging (Huang et al. 2015; Plank et al. 2016).

structure ideal for analyzing textual data and other sequentially-organized information.

From an implementation standpoint, I rely on the bidirectional LSTM neural network architecture proposed by Lample et al. (2016) and Ma and Hovy (2016).[10] Given a textual excerpt (e.g. a sentence or paragraph), this implementation generates a contextual representation for each token in the excerpt, which is used in a final tagging layer[11] to predict whether a given token is part of a named entity. Each token's contextual representation is constructed based on three sources of information:

1. Pre-trained embedding vectors for each word (here, drawn from GloVe, trained on the Google News corpus and described by Pennington et al. (2014));
2. Concatenated character embeddings for each character contained in the word (trained during model estimation); and
3. Embedding vectors and predicted tags for left- and right-adjacent terms.

This approach allows the final tagging layer to incorporate both simple and complex relationships. For example, incorporating character-specific information allows the model to easily learn that most named entities begin with a capital letter, while incorporating predicted tags for adjacent words allows the model to correctly tag multi-word named entities. Word embeddings, by contrast, incorporate more subtle information regarding word usage and semantic patterns, which can be used to identify words which are commonly contained in institution names of interest (e.g. "Secretary" or "Agency").

Using this set of examples, I trained the LSTM model described previously, and assessed its performance.[12] Two common performance metrics in the machine learning context are *precision* and *recall*. In the entity extraction setting, precision can be interpreted as positive predictive value, or the proportion of extracted entities that actually represent agency or institution names of interest. Similarly, recall can be interpreted as the true positive detection rate, or the proportion of total institution names of interest actually extracted by an

[10] As implemented in Tensorflow and Python by Guillaume Genthial. The implementation produced by Genthial is slightly different from the one outlined in the two papers I cite in-text; for details, see the accompanying documentation.

[11] Here, a conditional random field.

[12] Details regarding hyperparameter specification are given in Appendix A.

Table 1: Sample training example

Token	Tag
Funds	O
herein	O
appropriated	O
to	O
the	O
Department	B-MISC
of	I-MISC
Defense	I-MISC
for	O
construction	O
shall	O
be	O
available	O
for	O
hire	O
of	O
passenger	O
motor	O
vehicles	O
.	O

Sample output, formatted according to the CoNLL2003 format. Military Construction Act 1992 §102. For original text see the corresponding congress.gov page.

algorithm.[13]

Most algorithms perform better on one of these standards than the other. For example, dictionary methods offer a high *precision* (few false positives) but a low *recall* (many false negatives). By contrast, off-the-shelf, general-purpose named entity extraction methods offer the reverse, with a high *recall* (few false negatives) but a low *precision* (many false positives). A common evaluation metric that assesses both of these criteria simultaneously is the F_1 score, which is defined as the harmonic mean of precision and recall and offers a reasonable balance between both metrics.

How well does the LSTM approach perform by this standard? Since the purpose of using a machine learning approach for named entity recognition is to capture named entities not already known to the researcher, the most relevant (and stringent) performance test would be one in which we assess the model's ability to recover unseen entities not available during testing. In order to assess the model's performance in this scenario, I therefore conducted a cross-validated predictive accuracy study. In particular, I first randomly split my entity dictionary into five equally-sized groups. Beginning with the first group, I identified all sentences exclusively containing entities from the group in question, and used these sentences to form a held-out test set. I then trained a model using sentences containing entities from the remaining four groups, predicted values for the held-out test set, and used these predictions to calculate predictive accuracy and F1 score values. Finally, I repeated this process for each group in the dataset, and averaged the performance statistics to produce my final results.

Assessed in this fashion, the LSTM model I employ achieved a cross-validated F1 score of 0.758, with an overall accuracy of 0.965. These values are somewhat lower than others reported in the literature; for example, using a near-identical approach, Lample et al. (2016) report an F1 score of .904 on the standard CoNLL2003 named entity test dataset while Ma and Hovy (2016) report an F1 score of .912 and an overall accuracy of .976. However, the

[13]More precisely, $P = \frac{TP}{TP+FP}$ and $R = \frac{TP}{TP+FN}$, where P and R denote precision and recall, $T(P|N)$ denote the count of true positive/negative examples correctly classified and $F(P|N)$ denote the counts of true positive/negative incorrectly classified.

performance test I use in this paper is also noticeably more stringent than those used in other studies. In most studies of this kind, researchers assess performance by splitting *sentences* into training and test sets, rather than splitting *entities* into training and test sets. Any given entity therefore can (and usually will) occur in both the training and the test sets, creating a substantially simpler measurement ask. As a result, though the performance statistics I report leave some room for improvement, they represent a strong starting point from which to work.

3.3 Relation Extraction

Compared with entity extraction, relation extraction is a more difficult problem. Identifying a particular word or phrase as a named entity involves analyzing some data about that word or phrase (and perhaps its local context), and reaching a classification decision. By contrast, analyzing the *relationship* between two entities involves analyzing the entities, their local context, and any words or phrases which might encode information regarding their relationship. Worse still, relationships between actors are highly heterogeneous; for example, in a particular piece of legislation, two agencies might be given joint veto authority over the implementation of a policy program, or a court might be assigned to oversee and approve an agency's actions. Pre-specifying these relationship types and locating examples of each is a difficult problem, and is likely impractical when applied to a large corpus.

Fortunately, studying institutional complexity - my focus in this paper - does not require a detailed typology of relationship types. Institutional structures are complex to the extent that they include more actors in a particular policy area and create a denser network of decision-making structures that connect those actors. Though this definition might be enriched by a more nuanced definition of actor types or relationships, simply counting the number of actors and the relationships between them still reveals useful information on my key construct of interest. For both practical and theoretical purposes, I therefore focus on an abstract tie type, which is similar to that identified by Farhang and Yaver (2016). In particular, I define a tie between two actors as an instance in which two actors are *assigned to implement the same policy program.*

Luckily, drafting guidelines for American legislation - my primary application of interest in this paper - make these kinds of relationships relatively easy to identify. As noted in the drafting guide for the US Consolidated Code, the "basic unit" of every section of Code and legislation is the *section*.[14] Laws and Consolidated Code fragments are often further subdivided into ordered list elements of various types, but *sections* are intended to be stand-alone units that are roughly comparable in substantive scope. As a result, if we observe that two actors are co-mentioned in a section of a law, we can reasonably conclude that those two actors share authority over the policy area under consideration in that section. Without a sharper definition of the relationships under consideration, we cannot draw strong conclusions about the nature of the connections between these actors, but we can draw general conclusions about the basic implementation structure envisioned by the law in question.

3.4 Constructing Implementing Networks

Put together, the named entity and relation extraction approaches I describe offer a simple two-step implementing network construction procedure. First, we extract entities from each document using the pre-trained LSTM model described in §3.2, and discarded all named entities that were mentioned only once in their respective documents.[15] Second, we recombine each text into its constituent sections and drew an edge between any set of entities that co-occur in a given section, yielding an implementing network for each bill.

To illustrate this process, I provide two example outputs generated using this methodology in Figures 1 and 2. Beginning with the simpler case, the Enhanced Partnership with Pakistan Act of 2009[16] is a relatively straightforward foreign aid bill intended to provide military and developmental assistance to the Government of Pakistan. The law authorized the

[14] http://uscode.house.gov/detailed_guide.xhtml

[15] This heuristic is drawn from the natural language processing literature (see, e.g., Grimmer and Stewart 2013), and is useful in cases where typographic errors or other types of false positives are likely to be common.

[16] Public Law No. 111-73. For original text see corresponding congress.gov page.

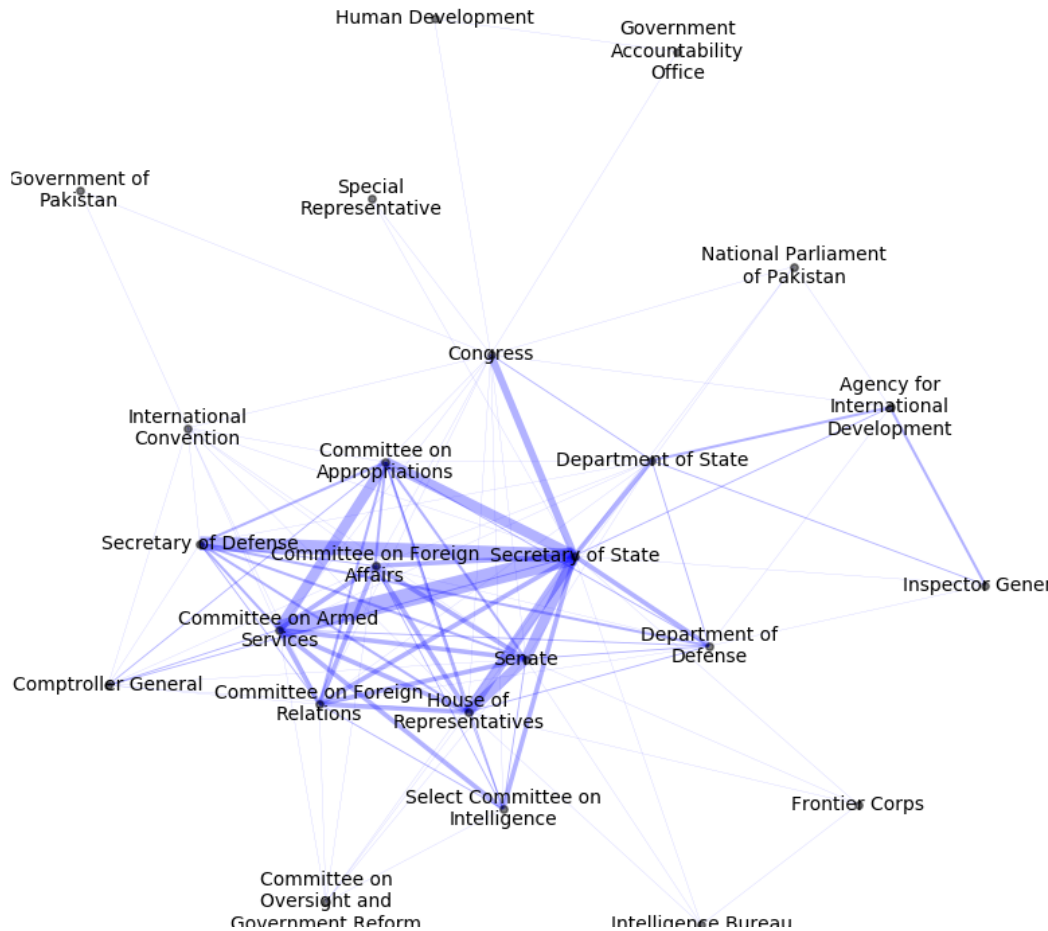
President to provide $1.5 billion in non-military aid from 2010-2014, and provided additional military aid conditional on a certification process implemented by the Secretary of State.[17] Unusually for a defense-oriented bill, the law gave the State Department substantial authority over defense-related aid allocations (Epstein and Kronstadt 2013). As shown in Figure 1, these features are clearly visible in the law's implementing network. The law contains a central cluster consisting of the Secretary of State, the Secretary of Defense, and several Congressional actors. Quantitative assessments of node importance reinforce this visual message; as measured by eigenvector centrality, the Secretary of State is the most central actor in this network (eigenvector centrality of 0.46), followed by the the Committee on Armed Services (0.37), and the House and Senate floors (0.33). These figures roughly track with qualitative summaries of the bill's content, lending this representation a substantial degree of face validity.

In contrast with the Enhanced Partnership with Pakistan Act, the American Recovery and Reinvestment Act (ARRA)[18] is both broader in scope and substantially more complex in its institutional organization. Briefly, the ARRA is a stimulus bill designed to bolster American economic performance following the 2007-2008 Financial Crisis. Since the ARRA covers so many policy areas and programs, a simple visual inspection of the network visualization is less informative. However, some patterns are readily apparent. Roughly speaking, the graph visualization algorithm I employ in Figure 2 places nodes with a greater number of connections between them in close proximity to one another. As a result, a simple visual inspection of the plot produced using this method reveals that the bill contains a central cluster corresponding to Congress (e.g. the House/Senate and various Congressional committees). Institutions covered by larger titles of the bill - such as Title VII (covering the Departments of Labor, Health and Human Services, and Education) and Title XII (Transportation and Housing and Urban Development) - form roughly coherent visual clusters, which are placed at the outer edges of the plot. This structure is reassuring; since the ARRA is roughly orga-

[17]Enhanced Partnership with Pakistan Act of 2009, §203.

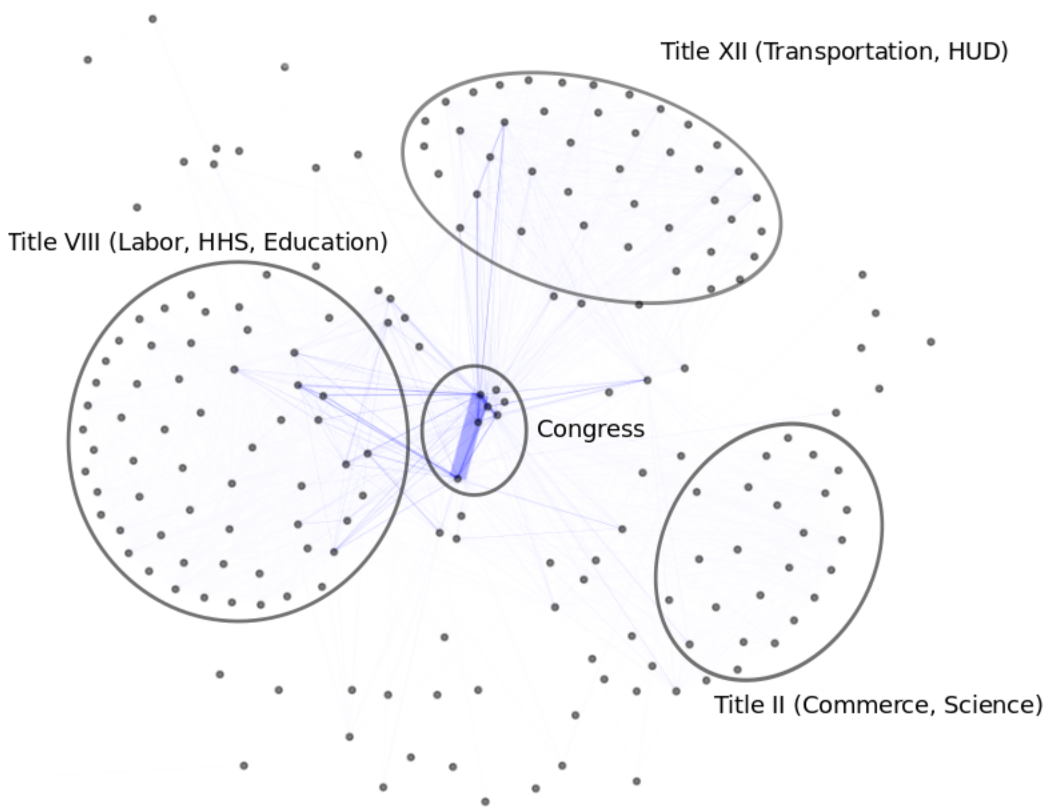
[18]Public Law No. 111-5. For original text see corresponding congress.gov page.

Figure 1: Implementing network, Enhanced Partnership with Pakistan Act of 2009.



Line density is approximately proportional to the number of ties between each node. Node placement is random, but is loosely related to node centrality.

Figure 2: Full implementing network for the American Recovery and Reinvestment Act of 2009.



Line density is approximately proportional to the number of ties between each node. Node placement is random, but is loosely related to node centrality.

nized by executive department, we should expect the bill's implementing network to display a roughly coherent set of clusters, which correspond to the major arms of the bill.

4 Institutional Complexity in American Law

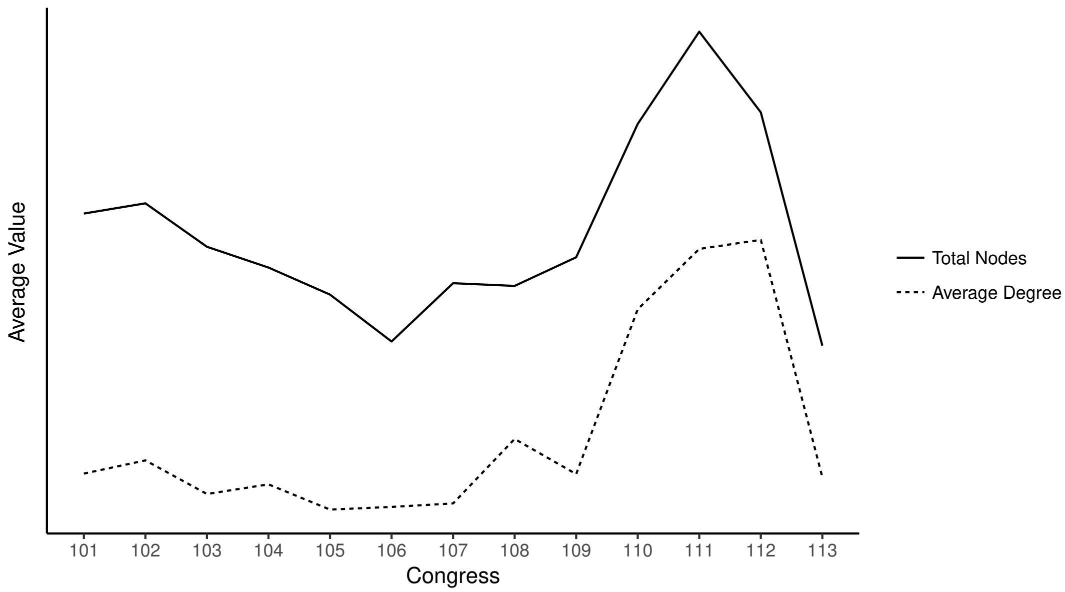
4.1 The Dependent Variable: Network Complexity

As I describe in §2.2, for the purposes of this paper I describe the implementing structures in a particular text as *complex* to the extent that the text grants a larger number of actors overlapping responsibility over execution of a particular policy program. This conceptualization offers at least two candidate measures of institutional complexity. First, we could simply count the *number of unique entities* identified in each law. This measure is simple and direct, but does not differentiate between implementing structures involving overlapping allocation of authority as opposed to a more "siloed" arrangement. An alternative is to measure the *average degree* of an implementing network, which we can interpret average number of instances in which a given actor is coinvolved with other actors in the implementation of a particular policy program.[19] Because this approach involves identifying both the correct set of actors named in a particular law and the set of relationships connecting those actors, it relies more heavily on identifying the correct set of textual features, but potentially offers greater information on the concept of interest.

Some basic descriptive information on both of these candidate measures for the American legislative dataset I describe in §3.1 are provided in Figure 3. Like Farhang and Yaver (2016) - which investigate a similar set of candidate measures - I find that the two complexity measures I propose are highly related. As shown in Figure 3, plotting the average fragmentation value for each measure and each session of Congress shows that these statistics essentially move in parallel throughout the period covered by my dataset. Because both measures are zero-inflated (27% zero values by total nodes; 41% by average degree) with

[19]Formally, the degree of a node in a real-valued network refers to the sum of all edge weights connected to that node. The average degree simply represents the average degree value over all nodes in the network

Figure 3: Comparison of two measures of legislative fragmentation.



substantial overdispersion (maximum values of of 566 total nodes and an average degree of 708), simple statistical measures of association like pairwise correlation values are misleading in this context. However, transforming each variable by a log-plus-one transformation[20] leads to a pairwise correlation value of 0.95. Because of this overlap, for modeling simplicity I focus on results generated using the node-count measure of my dependent variable for the remainder of this paper.

4.2 Predictor Variables

As I describe in §2, my key theoretical contentions are (1) previously observed relationships between preference disagreements and downstream design of legislation should be moderated by policy area and (2) formal institutional complexity should be driven primarily by issue salience and policy area than by preference-based factors. To operationalize executive-legislative preference disagreements, I use a simple binary indicator, which consists of a dummy variable indicating whether the same party controlled both the executive

[20] Defined as $ln_p(x) = ln(x+1)$.

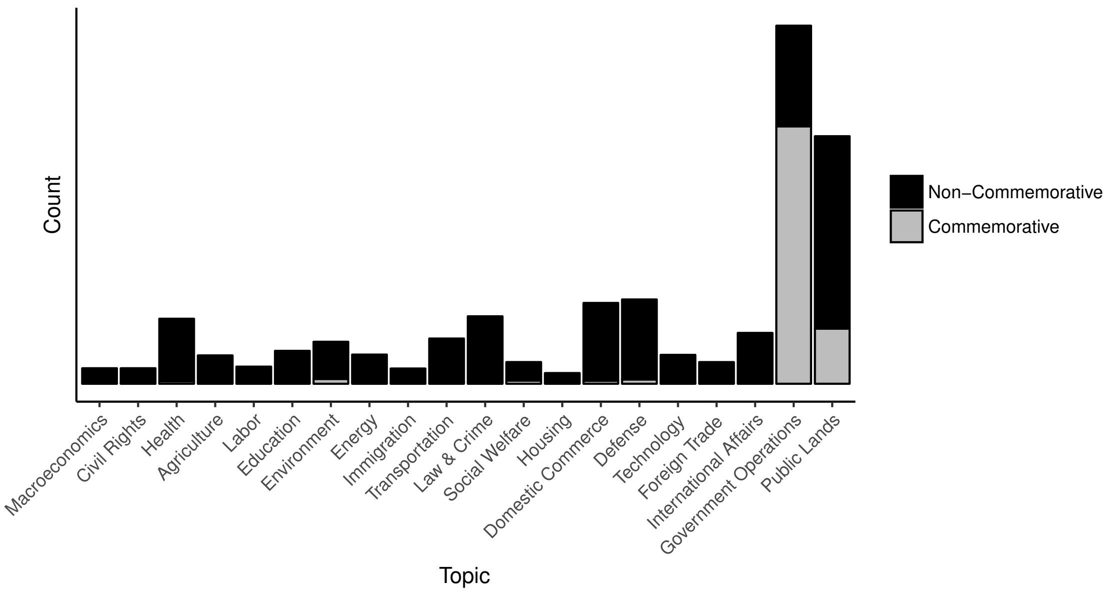
and legislative branches at the time that the bill was passed. As I outline in §2.1, most (but not all) existing studies have reported a negative relationship between unified government and institutional complexity. However, I argue that this relationship is likely to be more heterogeneous than previously reported, and attenuated for low-salience legislative texts.

Besides this basic partisanship variable, I also include two individual-level and two bill-level covariates. For the individual-level variables, I include covariates corresponding to the DW-NOMINATE score of the proposing member and an indicator variable denoting whether the proposing member was a part of the chamber majority. Generally speaking, more conservative members tend to be more skeptical of a strong administrative state; as a result, we might expect those members to be more willing to fragment implementing authority more frequently. By contrast, bills proposed by members of the chamber majority are likely to address higher-salience issue areas, making legislators more willing to craft complex and fragmented implementing structures.

For the bill-level variables, my primary theoretical quantities of interest are the *salience* and *policy area* of the bill in question. To operationalize policy area, I rely on the Congressional Bills Project's bill-level policy codes. This variable follows the Comparative Agendas Project's coding scheme covering some 20 major topic codes. As shown in Figure 4, all major topic codes are represented in this period, though some topics are substantially more common than others.

To operationalize salience, I use two separate variables. First, to measure the broader public visibility of a particular bill, I follow Volden and Wiseman (2014) and use a binary indicator denoting whether the bill in question was mentioned in the CQ Almanac's year-end summary of Congressional activity. Second, to measure salience within Congress I included a predictor corresponding to the square root of the number of cosponsors for each bill. Compared with my other variables, this measure is somewhat problematic; since important and time-sensitive bills often bypass the ordinary lawmaking process (Sinclair 2016), the

Figure 4: Enacted American legislation dataset, by policy area.



Topic names represent Comparative Agendas Project major topics. Commemorative bills are included here for completeness, but are not included in subsequent analyses.

authors of these bills may not have the time to gather a substantial number of cosponsors.[21] As an additional control, I therefore include a dummy variable indicating whether the bill in question was an annual appropriations bill or a defense authorization bill, which serves as a partial proxy for "must-pass" legislation.

Finally, to capture the interactive relationship between salience and legislative/executive preference disagreements I posit in §2, I include an interaction term between my `Unified_Government` and `CQ_Mention` variables. As I describe in §2, I expect the effects of preference disagreements on downstream design of legislation to be strongest on high-salience laws. As a result, the relationship between divided/unified government and fragmentation should be attenuated for low-salience bills. Put together, these expectations imply that the interaction term between the `unified` government and `CQ_Mention` variables should be small

[21] For example, contrast the Patient Protection and Affordable Care Act (Pub. L. 111-148) with the American Reinvestment and Recovery Act (Pub. L. 111-5). Though both bills were highly salient and the latter passed by substantially larger margins than the former, former bill had 40 cosponsors while the latter had only 9. This differential likely reflects the speed with which the bailout bill was enacted, compared with the more measured process for the Affordable Care Act.

Table 2: Descriptive statistics for predictor variables.

Variable	Mean	SD
Unified Government	0.31	0.46
DW-NOMINATE	0.1	0.5
Majority Sponsored	0.8	0.4
CQ Mention	0.3	0.5
$\sqrt{\text{Cosponsors}}$	2.6	3.0
Appropriation	0.03	0.18

on average, and should display the opposite sign from the base-level `CQ_Mention` coefficient within most policy areas.

5 Modeling

Since the unit of analysis in this model is the law, after dropping commemorative bills my final dataset contained 3494 observations, consisting of all enacted American legislation available via congress.gov and passed from approximately 1990-2014. To model patterns of institutional complexity within this dataset, I employed a Bayesian hierarchical negative binomial model with a hurdle component. We can loosely treat this model as a two-step regression, in which we first estimate a logistic regression to determine whether a given observation is zero or non-zero.[22] Within each part of the model, I estimate coefficients corresponding to each of the six variables I identify in Table 2.[23] For additional flexibility, I partially pool each coefficient estimate by policy area, allowing me to model policy area-specific effects for each variable.

My reasons for using this modeling structure follow directly from my theoretical expectations. The hurdle model I employ is an example of a mixture model, in which we treat the dependent variable as a mixture of two distinct probability distributions. In the context

[22]See Appendix B for details regarding model specification, estimation, convergence diagnostics, and posterior predictive checks.

[23]To prevent estimation issues, I exclude the `Appropriations` control from the logistic component of the model. See Appendix B for details.

of this project, I expect to encounter two types of bills: a “standard” type, which increases, decreases, or otherwise modifies the jurisdiction of one or more governmental actors, and a “non-administrative” type, which does not alter the jurisdiction of any actor. Examples of the latter type include “commemorative” bills[24] or bills which consist of technical amendments, corrections, or updates to other pieces of legislation.[25] Laws of this kind are likely to follow a different data-generating process than other bills contained in my dataset, with very few (usually zero) nodes and very few (usually zero) edges connecting any nodes that are present. The hurdle component of the model separates these kinds of bills from the dataset, allowing me to estimate separate coefficients for each of my predictors and each bill type.

Within both the hurdle and the count component of the model, I use a hierarchical prior structure to allow estimated coefficients to vary by policy area. Again, this choice follows follows directly from the expectations I outline in in §2. Hierarchical Bayesian models are particularly useful when we expect the coefficients associated with most predictor variables to interact with some underlying group structure. For the purposes of this paper, I expect the relationship between most of my predictor variables and my dependent variable to vary according to the policy area. In the count component of the model, since the `DW-NOMINATE` variable is scaled from 1 (most conservative) to -1 (most liberal) I expect the coefficients on that variable to be more positive on issues prioritized by conservative lawmakers (e.g. defense) and negative on issues prioritized by their more liberal counterparts (e.g. civil rights). I further expect the relationship between most of my predictor variables and my dependent variable to be larger in magnitude for higher-visibility policy, which reflects the interaction between policy/issue salience and other predictor variables I outline in §2. Using a hier-

[24]E.g. Pub. L. 102-262, “A bill to designate the United States Courthouse located at 111 South Wolcott in Casper, Wyoming, as the 'Ewing T. Kerr United States Courthouse'.” Recall that I filter bills identified as “commemorative” from the dataset before estimation; however, since the method used by the Congressional Bills Project to identify commemorative bills is heuristic and based on title keyword searches, some examples may slip through.

[25]E.g. Pub. L. 108-306, “To provide an additional temporary extension of programs under the Small Business Act and the Small Business Investment Act of 1958 through September 30, 2004, and for other purposes.” This bill simply extends authorization for existing provisions of the Small Business Investment Act of 1958, and therefore provides no modifications to existing administrative jurisdiction.

archical modeling structure allows for this kind of variability, offering additional analytical leverage.

6 Results

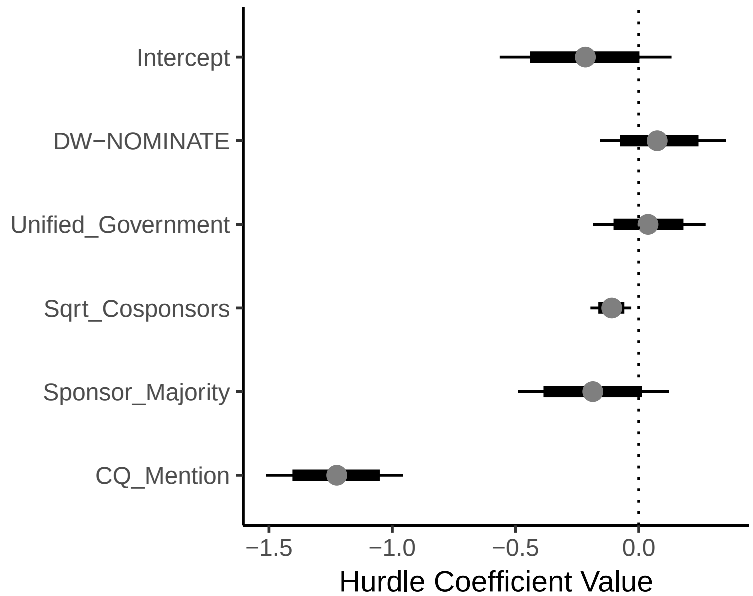
6.1 Hurdle Model

Beginning with the binary component of the model - which separates "administrative" from "non-administrative" bills - top-level posterior means and credible intervals for the hurdle model coefficients are given in Figure 5. The estimated posterior values for all coefficients match the expectations I outline above; bills with a greater number of cosponsors and bills that receive a mention in the CQ Almanac are both significantly less likely to be selected into the zero component of the model. The partisanship and unified government covariates have posterior credible intervals that cross zero for both top- and lower-level estimates, suggesting that these coefficients are largely unrelated to the dependent variable in this component of the model.

Since the hurdle component of the model is essentially equivalent to a standard logistic regression, we can easily transform coefficient estimates to more substantively meaningful values using standard techniques. Holding the coefficient on the `CQ_Mention` variable at its posterior mean, for example, suggests that receiving a mention in the CQ year-end almanac reduces the odds that a given observation will have a value of zero on the dependent variable by 71%. Policy-area specific estimates of this coefficient are roughly equivalent in in magnitude, suggesting that this variable's effect is approximately constant across policy areas. This finding is consistent with expectations; since most high-salience bills interact with some way with the administrative state, bills that receive press coverage are unlikely to be of the "non-administrative" type, no matter the policy area.

The coefficient on the cosponsor variable, by contrast, offers a good example of the payoff provided by the hierarchical coefficient structure I employ in this paper. As shown in

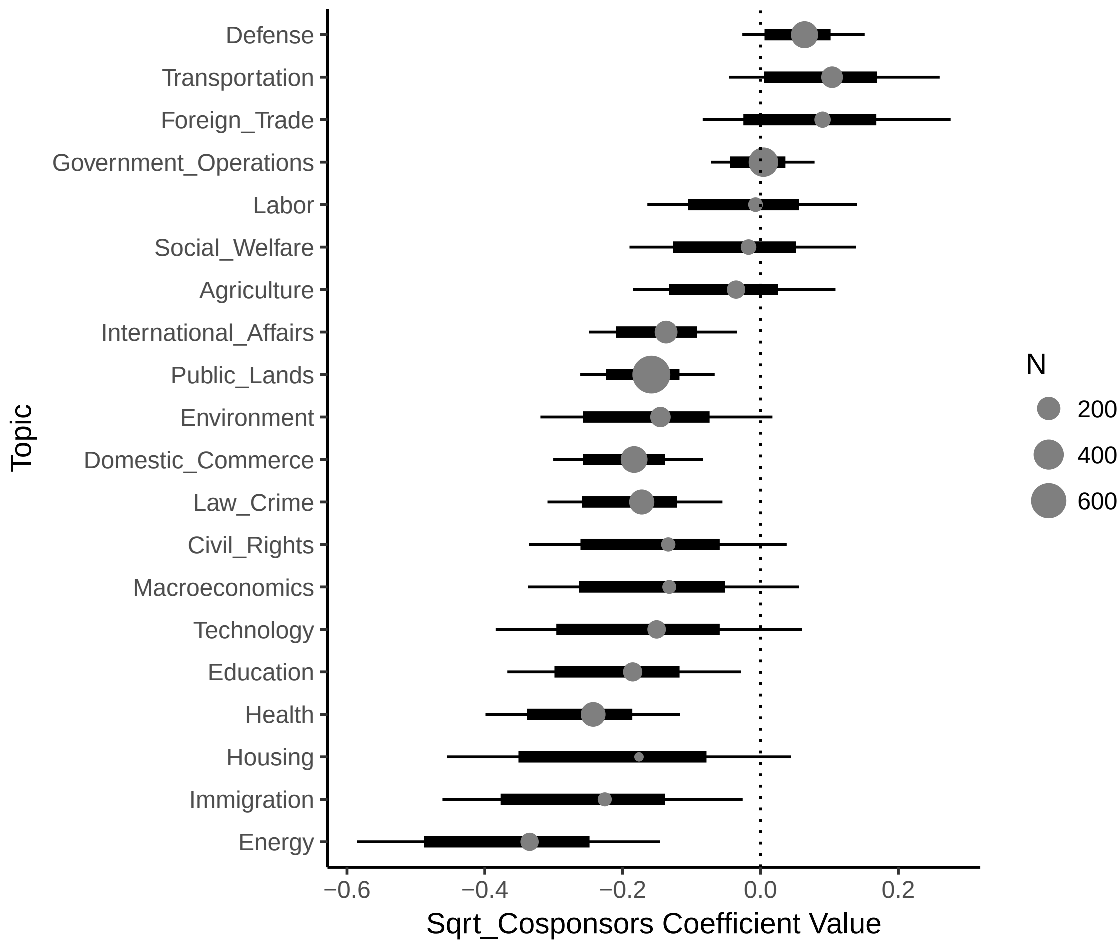
Figure 5: Top-level coefficients, hurdle model.



Dependent variable is a dummy variable indicating whether a given bill contains zero named entitites. Dots indicate posterior mean values. Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Positive estimates indicate that an increase in the given parameter value makes the dependent variable more likely to take on a value of zero.

Figure 5, the top-level posterior estimate of this coefficient is small but noticeably different from zero; a one-standard deviation decrease (≈ 3.0) in the `Sqrt_Cosponsors` produces an average 28% decrease in the odds that a given observation will have a value of zero on the dependent variable. However, as shown in Figure 6, estimates for this coefficient actually vary dramatically by policy area. For most bills, a greater number of cosponsors is associated with a small-to-moderate decrease in the probability that the dependent value will have a zero node count; however, for defense and transportation bills, a greater number of cosponsors actually *increases* the probability that a given bill will be of the "non-administrative" type. The reason for this difference is rooted in Congressional norms; since many of the most complex defense and transportation bills are "must-pass" funding measures, these bills usually bypass normal procedures, and attract few or no cosponsors (Sinclair 2016). As a result, bills in these domains which do attract cosponsors are more likely to contain few or no named entities.

Figure 6: Second-level `sqrt_cosponsors` coefficients, hurdle model.



Dependent variable is a dummy variable indicating whether a given bill contains zero named entitites. Dots indicate posterior mean values. Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Negative estimates indicate that an increase in the given parameter value makes the dependent variable less likely to take on a value of zero. Dot sizes scaled by the number of bills in each policy area.

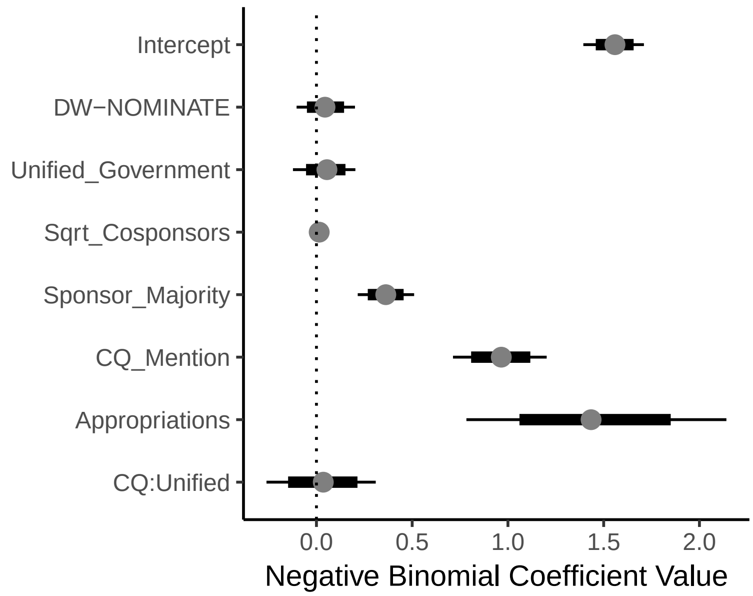
6.2 Count Model

Broadly, the count component of the model I present - which treats bill complexity as the dependent variable - can be interpreted similarly to the hurdle component. Since the count component uses a log-link, we can interpret the exponentiated coefficient estimates as having a multiplicative effect on the expected value of the dependent variable. For example, since posterior mean coefficient estimate for the `Sponsor_Majority` variable is ≈ 0.36, exponentiating this estimate yields a predicted $\approx 44\%$ increase in complexity when comparing majority-sponsored bills to their minority-sponsored counterparts. This result is highly intuitive; since members of the majority party control government, we should expect them to have both the capacity to enact more complex legislation and the inclination to pay closer attention to the internal workings of the administrative state. Using a similar procedure for the `Appropriations` variable yields a predicted $\approx 320\%$ increase in complexity. Though enormous, this latter estimate is also sensible. As I describe earlier in this section, appropriations bills are some of the highest-salience and most contentious bills in my dataset, which should lead us to expect these bills to be unusually complex.

As in the hurdle model, focusing on top-level coefficients can conceal substantial effect heterogeneity. For a clear example, consider the `CQ_Mention` and `Unified_Government` variables. Since I interact these two predicctors, we cannot consider their effects in isolation, which complicates interpretation. Fortunately, the Bayesian approach I use to estimate the model in this section enables a simple solution. Starting with the `CQ_Mention` variable, to generate the estimated coefficient when `Unified_Government` $= 1$, we can add the posterior draws from each iteration for the `CQ_Mention` and `CQ_Mention:Unified` interaction variables, and use the results to produce posterior mean and credible intervals for this scenario. To generate estimates when `Unified_Government` $= 0$, we can simply use the raw posterior draws for the `CQ_Mention` coefficient.

The results of this procedure are shown in Figure 8. Though this variable's estimated effect is large and positive in for most policy areas, the scale of its effect varies dramatically.

Figure 7: Top-level coefficients for the count model.

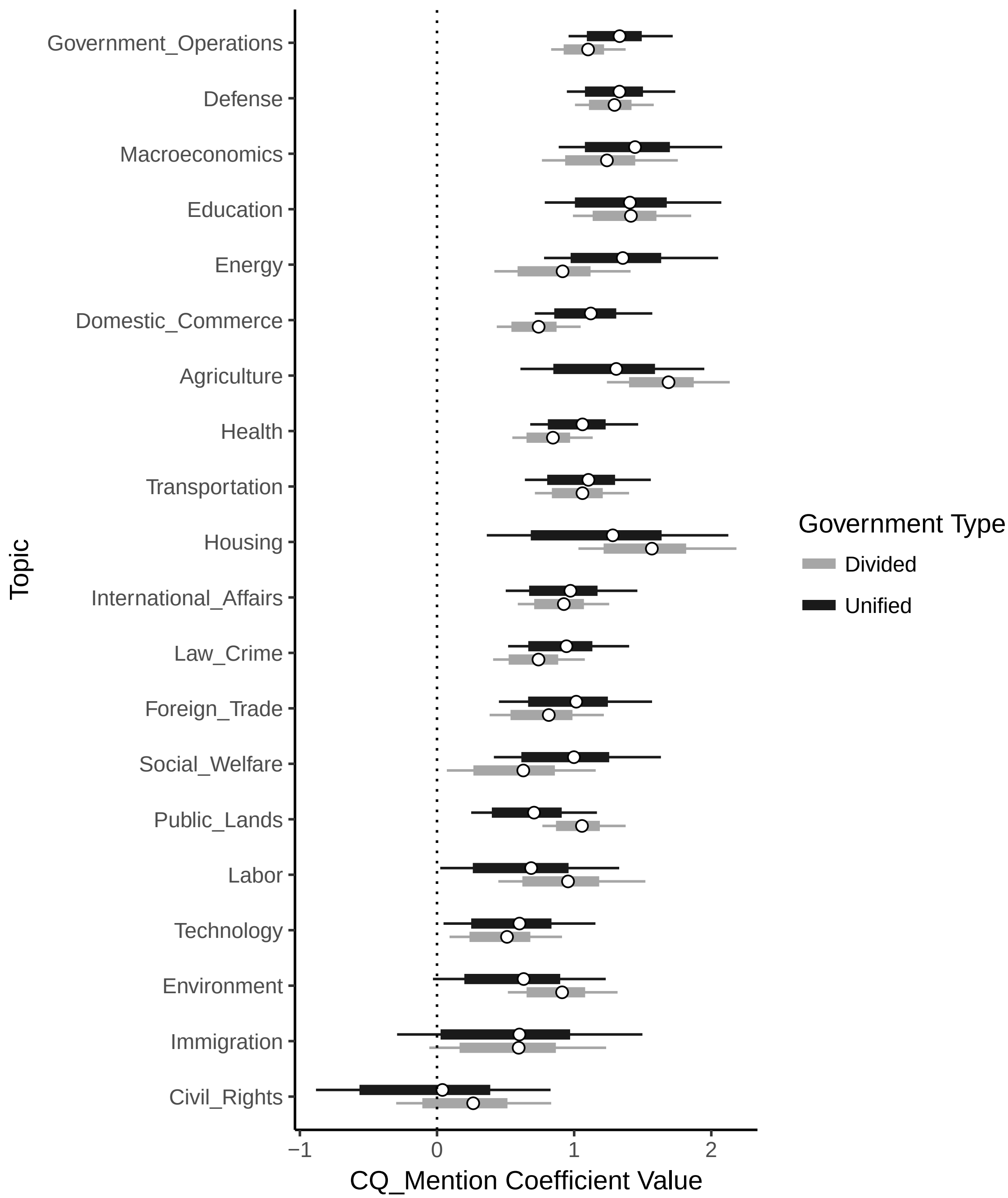


Dependent variable is a count of unique named entities contained in a bill. Dots indicate posterior mean values. Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Positive estimates indicate that an increase in a given coefficient increases fragmentation.

Under unified government, being mentioned in CQ's year-end almanac yields a predicted 300-400% increase in fragmentation in policy areas like government operations, defense, and macroeconomics. By contrast, public salience affects bills addressing civil rights, immigration, and environment much more modestly. This broad pattern remains similar under divided government, though rankings across policy areas are somewhat shifted.

We can use a similar strategy to investigate effect heterogeneity in the `Unified_ Government` variable. As I predict in §2, visual inspection of Figure 9 suggests that the relationship between partisanship and downstream design of legislation is attenuated for non-salient bills. In other words, bills that receive little or no public attention possess essentially equivalent implementing structures when passed under unified or divided government. By contrast, bills that receive substantial public attention vary more noticeably when passed under divided and unified government, with most coefficient estimates being larger in absolute value for salient bills. A Bayesian posterior probability calculation verifies this visual inspection, suggesting that - averaged across all policy areas - there is an estimated 78% posterior prob-

Figure 8: Second-level `CQ_mention` coefficients, count model. Estimates for both unified and divided government scenarios included.



Dependent variable is a count of unique named entities contained in a bill. Dots indicate posterior mean values. Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Positive estimates indicate that an increase in the coefficient increases the bill's estimated fragmentation value.

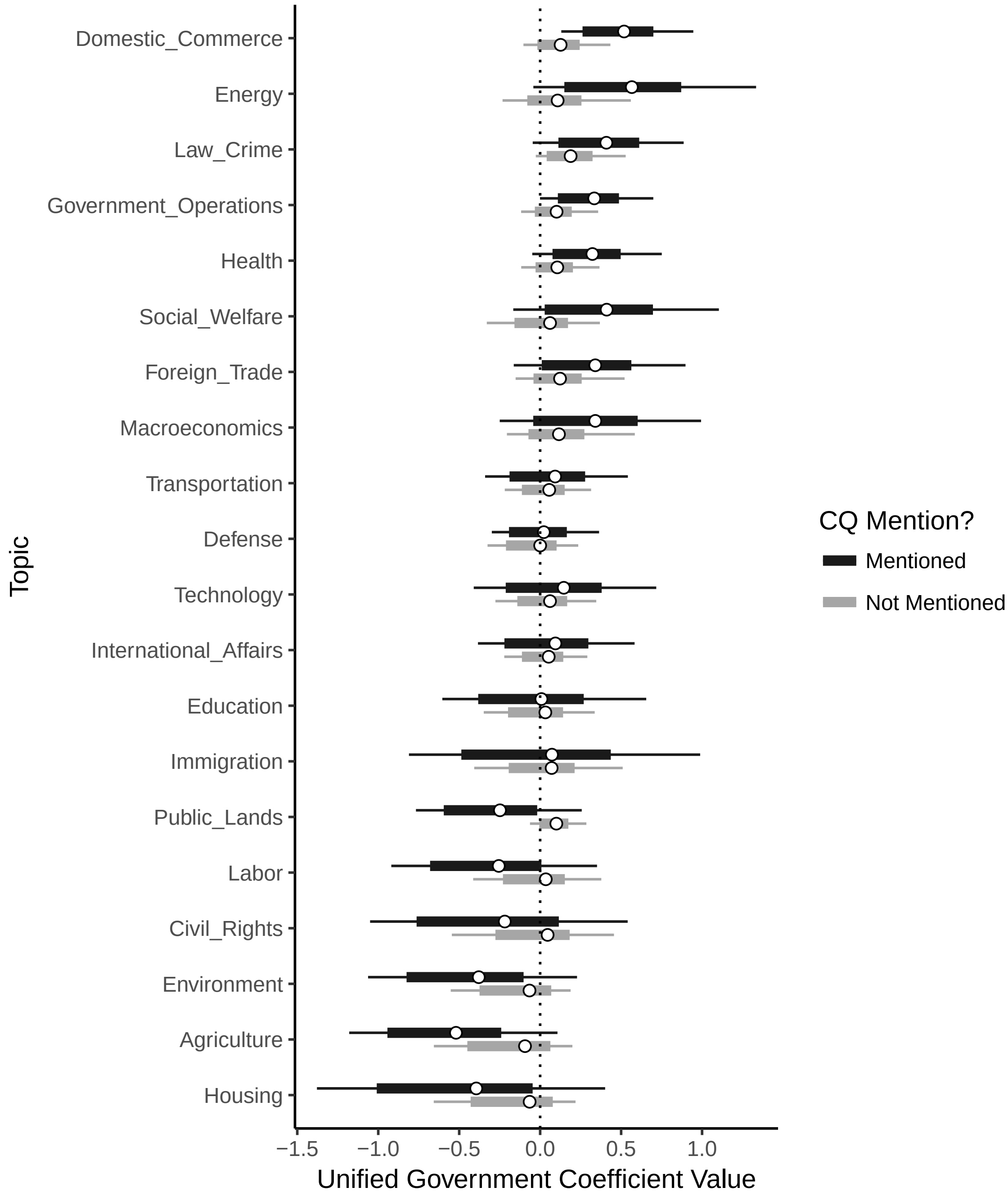
ability that a given coefficient estimate will be larger in absolute value for high-salience bills than low-salience bills. This attenuation pattern is largest for policy areas like domestic commerce, energy, government operations, and agriculture, but is noticeable in most policy areas in the dataset.

These findings complicate results from the existing literature, and offer support for the hypotheses I present earlier in this paper. As predicted, the relationship between preference-related variables like divided/unified government and downstream allocation of authority decisions is heterogeneous and weak on average. Reassuringly, in the policy areas where the estimated relationship between divided/unified government and downstream allocation of authority is strongest, the results I report line up with those from at least some existing studies. For example, Huber and Shipan (2002) and find a positive relationship between discretion (an idea related to fragmentation) and unified government in health-related legislation, which matches the coefficient I report for high-salience bills. However, because the estimates across policy areas are so heterogeneous, we should be skeptical regarding the applicability of this pattern to other policy settings and to lower-salience pieces of legislation.

7 Conclusion

Overall, the results I present reinforce the arguments I present at the outset of this paper. At least in the modern American context, formal institutional design is primarily driven by *policy* rather than *political* considerations. Under both divided and unified government, politically salient bills are more complex than their lower-salience counterparts. However, this relationship also varies substantially by policy area. High- and low-salience bills in inherently "complex" policy areas involving large disbursements of public funds - such as macroeconomics and defense - differ much more dramatically than their counterparts in areas involving lower levels of public spending. By contrast, in nearly all policy areas, bills passed under divided and unified government differ little in terms of the complexity of their

Figure 9: Lower-level `Unified_Government` coefficients, count model.



Dependent variable is a count of unique named entities contained in a bill. Dots indicate posterior mean values. Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Positive estimates indicate that an increase in the coefficient increases the bill's estimated fragmentation value.

institutional structures.

These results emphasize the importance of both the theoretical ideas and the measurement techniques I introduce in this paper. In existing work on legislative fragmentation, measurement constraints have forced authors to restrict their attention to single policy areas or to "significant" policy areas. However, as I show, both of these factors substantially affect the design of legislation, both on their own and by structuring the relationships of other predictors. Without the measurement techniques I introduce, these findings would not have been possible to produce, emphasizing the importance of scalable, broadly applicable measurement techniques for applied work.

The results in this project offer a number of directions for future research. Within the domain of American legislation, comparing enacted legislation to proposed but unpassed bills offers opportunities understand the sources of institutional complexity in a direct fashion, both at an individual and an institutional level. More broadly, though the application in this paper is restricted to American national-level legislation, the framework and tools I present are broadly applicable to legal documents from other contexts. As a result, the measurement approach I propose offers opportunities to study lawmaking patterns in different social and political contexts and under different systems of government.

References

Bendor, J. and Meirowitz, A. (2004). Spatial models of delegation. *American Political Science Review*, 98(02):293–310.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2016). Stan: A probabilistic programming language. *J Stat Softw*.

de Figueiredo Jr, R. J. (2002). Electoral competition, political uncertainty, and policy insulation. *American Political Science Review*, pages 321–333.

Epstein, D. and O'Halloran, S. (1999). *Delegating powers: a transaction cost politics approach to policy making under separate powers*. Cambridge University Press, New York.

Epstein, S. B. and Kronstadt, K. A. (2013). Pakistan: Us foreign assistance. *Current Politics and Economics of the Middle East*, 4(3):575.

Farhang, S. and Yaver, M. (2016). Divided government and the fragmentation of american law. *American Journal of Political Science*, 60(2):401–417.

Franchino, F. (2004). Delegating powers in the european community. *British Journal of Political Science*, 34(02):269–293.

Franchino, F. (2007). *The powers of the union: delegation in the EU*. Cambridge University Press, New York.

Franzosi, R., De Fazio, G., and Vicari, S. (2012). Ways of measuring agency: an application of quantitative narrative analysis to lynchings in georgia (1875–1930). *Sociological Methodology*, 42(1):1–42.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*. Taylor & Francis, Boca Raton, FL, 3 edition.

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

Huang, Z., Xu, W., and Yu, K. (2015). Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

Huber, J. D. and Shipan, C. R. (2002). *Deliberate discretion? the institutional foundations of bureaucratic autonomy*. Cambridge University Press, New York.

Huber, J. D., Shipan, C. R., and Pfahler, M. (2001). Legislatures and statutory control of bureaucracy. *American Journal of Political Science*, pages 330–345.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*.

Leetaru, K. and Schrodt, P. A. (2013). Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, volume 2, pages 1–49. Citeseer.

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

Ma, X. and Hovy, E. (2016). End-to-end sequence labeling via bi-directional lstm-cnns-crf. *arXiv preprint arXiv:1603.01354*.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., and McClosky, D. (2014). The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60.

Mayhew, D. R. (1991). *Divided we govern.* Yale University.

McCubbins, M. D. and Schwartz, T. (1984). Congressional oversight overlooked: police patrols versus fire alarms. *American Journal of Political Science*, pages 165–179.

Moe, T. M. (1990a). Political institutions: the neglected side of the story. *Journal of Law, Economics, & Organization*, 6:213–253.

Moe, T. M. (1990b). The politics of structural choice: toward a theory of public bureaucracy. In Williamson, O., editor, *Organization theory: from Chester Barnard to the present and beyond*, pages 116–153. Oxford University Press, New York.

Moe, T. M. (2012). Delegation, control, and the study of public bureaucracy. In *The Forum*, volume 10.

Moe, T. M. and Caldwell, M. (1994). The institutional foundations of democratic government: a comparison of presidential and parliamentary systems. *Journal of Institutional and Theoretical Economics*, 150(1):171–195.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Plank, B., Søgaard, A., and Goldberg, Y. (2016). Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss. *arXiv preprint arXiv:1604.05529*.

Sinclair, B. (2016). *Unorthodox lawmaking: New legislative processes in the US Congress.* CQ Press.

Sundermeyer, M., Schlüter, R., and Ney, H. (2012). Lstm neural networks for language modeling. In *Thirteenth Annual Conference of the International Speech Communication Association*.

Volden, C. and Wiseman, A. E. (2014). *Legislative Effectiveness in the United States Congress: The Lawmakers*. Cambridge University Press.

A Statutory Text Parsing Details

A.1 Header Regular Expressions

Table 3 gives the set of regular expressions used as inputs to the constitute_tools parser, which I use to parse the American legislative text database I use in this paper. Note that the list of regular expressions given in this table does not fully capture the set of organizational levels present in American legal language. For example, nearly all major American legislation contains "Title" or variants, which are not captured by this list.

Table 3: Regular expressions used to parse American legislative texts

Regular Expression	Sample Plain-Text Match
`(SECTION|SEC\.)\s*\.?\s*(<< NOTE: [0-9]+ USC [-0-9a-z]+\.?\s*')]`	SECTION 101; SEC. 446a
`(note)?\.?>>)?\s*[0-9]+\.\s*`	344.
`\([a-z]\)`	(a)
`\([0-9]+\)`	(34)
`\([A-Z]\)`	(C)

My rationale for excluding some kinds of organizational headers from this list is straightforward. Like many legal corpora, American legislative texts are not written in an entirely consistent fashion. As a result, attempting to capture all organizational headers present across the whole corpus I examine would be highly labor-intensive. Moreover, attempting to extract an organizational header without fully understanding the range of variation present in the corpus can actually do more harm than good; since regular expressions are so flexible, including an overly broad set of regular expressions intended to capture a missing level can potentially consume legally significant language, leading to missing information. By contrast, excluding an organizational header entirely simply introduces a few extraneous words into the set of language under consideration, especially for rarer headers like "Title".

The selection of regular expressions that I do include in 3 are included for two reasons. First, as I note in-text, the Office of Law Revision Counsel stipulates that "Sections" of American legislation should be comparable in their substantive scope, which is a structural standard on which I heavily rely throughout my analysis. Segmenting sections is therefore the most critical part of the parsing and text-cleaning task, which is why the regular expression corresponding to "Section" in Table 3 captures so many special cases and variations. Second, the other headers I include are ubiquitous in nearly every document in the dataset; as a result, excluding them from the list would introduce substantial extraneous language.

A.2 Sample Parsed Text

Table 4: Sample parsed document

Title	Text
SEC 416	FOREIGN STUDENT MONITORING PROGRAM.
(a)	Full «NOTE: 8 USC 1372 note.» Implementation and Expansion of Foreign Student Visa Monitoring Program Required.–The Attorney General, in consultation with the Secretary of State, shall fully implement and expand the program established by section 641(a) of the Illegal Immigration Reform and Immigrant Responsibility Act of 1996 (8 U.S.C. 1372(a)).
(b)	Integration «NOTE: 8 USC 1372 note.» With Port of Entry Information.–For each alien with respect to whom information is collected under section 641 of the Illegal Immigration Reform and Immigrant Responsibility Act of 1996 (8 U.S.C. 1372), the Attorney General, in consultation with the Secretary of State, shall include information on the date of entry and port of entry.

USA PATRIOT Act §416(a-b). For original text see the corresponding congress.gov page.

A.3 LSTM Parameter Specification

As described in §3, I use an LSTM to extract named entities from legislative texts. For the hidden character and word embedding layers, I used a layer sizes of 100 and 300 nodes, respectively. As mentioned in-text, rather than training word embeddings directly I used pre-trained embeddings drawn from Pennington et al. (2014)'s GloVe dataset. Like virtually all neural network applications, I trained this model using stochastic gradient descent.[26] I trained the model for 5 epochs, using 90% of pre-identified named entities for training and 10% as a held-out test set. To avoid overfitting, I use a dropout rate of 0.5 and a batch size of 20, with a learning rate of 0.015, a learning rate decay of 0.05, and a gradient clipping value of 5.0.

[26]Specifically, an ADAM optimizer. See Kingma and Ba (2014) for details.

B Bayesian Model Details

B.1 Specification

For the model I describe in §6, I use the following likelihood function:

$$p(y_i|\rho_i, \theta_i, \phi) = \begin{cases} \rho_i & \text{if } y_i = 0 \\ (1-\rho_i)\frac{NB(y_i|\theta_i,\phi)}{1-NB_{CDF}(0|\theta_i,\phi)} & \text{if } y_i = 1 \end{cases}$$

With y_i the node count for the i^{th} observation, NB the negative binomial density, and NB_{CDF} the negative binomial CDF.[27] We can (very) loosely treat this model as a two-step regression, in which we first estimate a logistic regression to determine whether a given observation is zero or non-zero. For the set of observations selected into the non-zero component, we then estimate a negative binomial regression, in which the likelihood for the model is truncated at one.

To incorporate covariates into this model, I define:

$$\rho_i = logit(X_i'\gamma_{z_i})$$

$$\theta_i = log(X_i'\beta_{z_i})$$

with X a $N \times (K+1)$ matrix of predictors as defined in Table 2, γ a $(J+1) \times M$ and β a $(K+1) \times M$ matrix of regression coefficients, and z_i an auxiliary $N \times 1$ vector of group assignments for each observation. This structure defines a standard hierarchical model, in which we separately estimate regression coefficients for both the logistic and negative binomial components of the mixture for each group. Here, I use the *policy area* of each bill as the group, yielding $M = 20$ unique groups for the dataset. I include all predictor

[27] As defined using the location-scale negative binomial parameterization in Carpenter et al. (2016), in which $E(y_i) = \theta_i$ and $Var(y_i) = \theta_i + \frac{\theta_i^2}{\phi}$. Since ϕ is not subscripted, this structure implies a conditional constant variance assumption, in which I assume the variance of each observation to be constant for a given value of θ_i.

variables listed in §4.2.2 as well as an interaction between the `Unified_Government` and `CQ_Mention` variables in the count component of my model (yielding $K = 7$). For the count component of the model, I exclude the `Appropriations` variable and the interaction term to avoid separability-related estimation issues (yielding $J = 5$).

Fitting this model with independent coefficients involves estimating $M((K+1)+(J+1))$ regression coefficients, which may lead to high-variance coefficient estimates or problems with model convergence. To stabilize estimates, I therefore place a shared prior on each set of group-level coefficient estimates:

$$\mu_\gamma \sim MVN(\gamma, \Sigma_\gamma)$$

$$\mu_\beta \sim MVN(\beta, \Sigma_\beta)$$

With μ_γ and μ_β each a $(K + 1)$ vector of top-level regression coefficients and Σ_γ a $(J + 1) \times (J + 1)$ and Σ_β a $(K + 1) \times (K + 1)$ variance-covariance matrix. This prior structure "partially pools" coefficient estimates for each predictor, allowing the data to inform the model regarding the extent to which policy area-specific coefficients for each variable should be allowed to vary. In cases where we should expect a given coefficient's value to vary substantially across policy area, this prior structure allows coefficients to vary appropriately. However, in cases where the effect of a given coefficient is more uniform, the partial pooling structure prior structure allows coefficient posteriors to borrow precision from one another, reducing variance in these estimates.

To complete the model, I place a vague half-normal prior on the negative binomial scale parameter $\phi \sim N_T(0, 5)$. I then place priors on the variance-covariance matrices Σ_γ and Σ_β using the strategy suggested by Gelman et al. (2014). I first define auxiliary variables ν_β, ν_γ, Ω_β, and Ω_γ using the general form $\Sigma = diag(\nu)\Omega diag(\nu)$. This decomposition eases estimating by allowing me to place separate priors on the location and scale of the variance-covariance matrix for each set of coefficients. For numerical stability, I further decompose

Ω using a Cholesky factorization such that $\Omega = LL'$, and place the following priors on the auxiliary variables:

$$\nu_\gamma, \nu_\beta \sim N_T(0, 10)$$
$$L_\gamma, L_\beta \sim LKJ(1)$$

With N_T a half-normal prior, and LKJ denoting the Lewandowski et al. (2009) correlation matrix distribution. $LKJ(1)$ reduces to an identity distribution over correlation matrices, which causes this prior to represent a flat prior over coefficient correlation. The prior on ν was selected to represent a vague but mildly informative prior, indicating a slight preference towards coefficient estimates that are smaller in absolute value. In most situations, priors of this kind aid numerical stability during estimation and improve posterior predictive performance.

B.2 Estimation and Posterior Predictive Checks

To fit the model, I used the Stan programming language (Carpenter et al. 2016). I ran four chains, with 1000 warmup iterations and 3000 post-warmup iterations in each chain. I initialized all parameter values at 0, and used a maximum treedepth of 15 and an `adapt_delta` value of 0.98. Initial experiments suggested that the default maximum treedepth (10) was sometimes exceeded and a small number of divergent transitions were sometimes encountered with the default adaptation-phase acceptance probability. Increasing these parameters eliminated these problems. Visual plots suggested good mixing across chains, with $1 \leq \hat{R} \leq 1.01$ for all parameters and $n_{eff} \geq 1000$ for all parameters.[28]

Following Gelman et al. (2014), in Figure 10 I visually assess model fit using posterior predictive checks. In each plot, I provide the observed density of the node count dependent variable, overlaid on density plots for 400 simulated dependent variable datasets based on randomly-selected post-warmup posterior parameter draws. As shown in the left-hand panel, across the whole dataset the model fit is excellent. Zooming in on smaller values (where most posterior density is located) reveals that the model slightly under-fits at very small values of the dependent variable ($y \leq 10$). Even in this range, however, model fit remains acceptable.

[28]With $\hat{R}$ a diagnostic quantifying the consistency of an ensemble of Markov chains, and n_{eff} a rough effective sample size calculation (Gelman et al. 2014).

Figure 10: Posterior predictive plots for the node count dependent variable