Power in Text: Extracting Institutional Relationships from Natural Language

Robert Shaffer
University of Texas, Austin
rbshaffer@utexas.edu

June 29, 2018

Abstract

How do legislators allocate policy-making authority? Distribution-of-power decisions inevitably involve a trade-off between efficiency and accountability. Fragmenting decision-making authority between several actors increases oversight opportunities, but limits the effectiveness of individual actors. At least in the legal context, these decisions are usually articulated in written texts. Unfortunately, extracting this information from text is difficult, leading major studies to restrict their scope to a single policy area or to a small set of significant legislation.

In this project, I address these limitations in two respects. Methodologically, I propose a neural network-based approach designed to extract power relationships from legal texts in a scalable, valid fashion. I then apply this approach to study fragmentation of implementing authority in American legislation. Overall, I find that issue salience is strongly and positively related to fragmentation on average, but that the magnitude of this effect is conditional on policy area. In contrast with existing work, I find that ideological differences between the executive and the legislator are relatively weak predictors of fragmentation patterns, highlighting both the importance of the tools I develop and the need for scalable measurement techniques in political science.

1 Introduction

How do legislators allocate policy-making 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. To take a pair of (seemingly) extreme examples from American politics, 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 as an efficiency/accountability tradeoff. Particularly in the American context, researchers have formalized this intuition through versions of the "ally principle" (Epstein and O'Halloran 1999; Huber and Shipan 2002; Bendor and Meirowitz 2004; Franchino 2007; Gailmard and Patty 2012; Farhang and Yaver 2016). When the executive's interests are aligned with those of the legislature (and are likely to remain so in the future), these authors argue, legislators are more willing to pass simple "framework"-style legislation. By contrast, when legislative and executive interests diverge (or might do so in the future), legislators tend to favor complex institutional structures, which provide greater oversight opportunities and fragment decision-making authority among many actors.

Beyond these basic political factors, some writers have also posited that the design of a given legal document should be affected by characteristics of the *issues* and *policy areas* that document addresses (e.g. Epstein and O'Halloran 1999; Bendor and Meirowitz 2004). For example, we should expect lawmakers to be more willing to design complex decision-making structures when addressing issues that are *salient* for their constituents, or when addressing policy areas in which lawmakers possess some preexisting expertise. Unfortunately, while issue and policy area variables have been discussed extensively in the theoretical literature on formal allocation of authority, ideas like these are largely ignored in empirically-oriented scholarship. Largely, this limitation result 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.

In this paper, I address these limitations in three parts. First, from a theoretical standpoint, I argue that the process of constructing a legal document is best viewed in terms of a cognitive cost-benefit analysis. This framework highlights the importance of *issue salience* and *policy area*, which have largely been ignored by the existing institutional design literature. Second, 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 test on an original dataset of American legislation. Finally, I use this measurement approach to study fragmentation of authority in an original dataset consisting of all enacted American laws passed from 1990-2016. As predicted, I find that *issue salience* and *policy area* are by far the most important predictors of institutional fragmentation, highlighting the importance of the measurement tools I develop.

2 Lawmaking and the Costs of Fragmentation

2.1 Cognitive Microfoundations

When a policymaker writes or modifies a formal policy document - such as a constitution, law, or administrative regulation - that policymaker must choose between a range of possible implementing models. 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, delegates most authority to one or a few implementing actors. The choice between these two models is frequently consequential;

by dividing policymaking authority between several actors, lawmakers reduce implementing efficiency and limit implementor 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.

Most scholars have framed these kinds of institutional design choices as an efficiency/accountability tradeoff. As Moe (1990a; 1990b; 2012) and Moe and Caldwell (1994) argue, complex implementing structures are not *ex ante* desirable. Complicated power-sharing arrangements that fragment implementing authority across an array of actors dramatically curtail administrative flexibility, reducing implementer responsiveness and promoting policy gridlock. Moreover, detailed authority structures are not costless to create. If a policymaker wants to design a complex policy structure with the intention of accomplishing a policy objective, he or she needs to devote substantial time and attention to the process, which usually involves a substantial quantity of research, expert consultation, and experimentation. Thus, at least in the abstract, policymakers should prefer "framework" legislation that offers one or a few implementing actors substantial discretionary authority.

Here, Moe's "politics of structural choice" intervenes. Whenever policymaker and implementor possess different substantive preferences, formal mechanisms of bureaucratic control - institutional or otherwise - become more attractive as a policy strategy, offering protection for policymakers against implementer malfeasance (McCubbins 1985). Issue salience and policy complexity provide a similar incentive structure. From both a rational-choice and a bounded-rationality perspective, legislators possess greater incentives to restrict executive authority in high-salience, low-complexity policy areas (e.g. Bendor and Meirowitz 2004; Baumgartner et al. 2009). In these domains, the electoral and policy payoffs for creating highly restrictive institutional arrangements which "lock in" legislator preferences are high (because their impacts are visible) and the costs are low (because the cognitive costs involved in researching and designing complex systems in these areas are small).

2.2 Measurement and Selection Bias in Existing Work

Unfortunately, though theoretical predictions in this literature are deep and well-developed, empirical verification of these ideas is relatively shallow. The reason for these 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 naturally forced scholars to limit the scope of their work, in at least two important ways. Most prominently, little empirical work has 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 the face the measurement challenges endemic in this literature, self-imposed limitations like these are a natural response. Unfortunately, work in American politics and public policy gives us good reason to suspect that both policy area and issue salience will have important impacts on downstream choices regarding allocation of authority. To reiterate, complicated institutional structures are cognitively demanding to construct. As a result, we should expect lawmakers to be most willing to pay these costs when gap between executive and leigslative policy preferences is large *and* when the policy area(s) and issue(s) in question are relatively simple and high-salience, respectively. Focusing on single policy areas or historically "significant" legislation may therefore overstate the relationship between preference disagreements and downstream design of legislation, at least for day-to-day lawmaking activities.

2.3 Hypotheses

To fix intuition, before proceeding I provide a formalized statement of the hypotheses I describe in the previous two sections. These hypotheses can be summarized in four basic statements.

H1: All else equal, legislators should create simpler authority structures in complex policy areas outside their core competences than when addressing more familiar and straightforward policy problems.

This idea follows directly from the bounded rationality literature. If a policymaker possesses less experience working in a particular policy area, creating legislation designed to address problems within that area will require greater cognitive effort. Since policy experience varies by individual, the "cognitive load" Sweller (1994); Paas et al. (2003); Sweller (2010) imposed by a particular legislative proposal will also vary, but we might reasonably expect the average cognitive load for a member of Congress to be lower in familiar, high-profile policy areas like finance or criminal policy than in more technical ones like space and science regulation.[1]

H2: All else equal, legislators should create simpler authority structures when addressing low-attention policy problems than when examining their higher-salience counterparts.

The individual-level foundations for this hypothesis are very similar to those I offer for **H1**. On average, low-salience policy issues attract little legislative attention, since legislators are usually disinclined to exert the cognitive effort required to create complex decision-making structures without constituent pressure. As a result, when addressing low-attention policy issues legislators should be more likely to delegate decision-making authority to one or

[1] Epstein and O'Halloran (1999) provide some support for this idea, noting that legislators tend to delegate more on high-complexity policies (e.g. space policy).

a few implementing actors, instead of creating a complex and interconnected decision-making structure.

H3: If the executive and legislative actors possess similar preferences, legislative outputs should contain simple decision-making structures. By contrast, if legislative and executive actors have different preferences, decision-making structures may be either simple or complex.

As described in previous sections, the relationship between legislative/executive preference conflicts and downstream allocation of authority is complex. However, as Volden (2002, 112) notes, "both the executive and the legislature do indeed have an interest in increasing bureaucratic discretion when their preferences align".[2] This idea - which is further developed by Bendor and Meirowitz (2004) - suggests that set of desirable decision-making structures is relatively small when executive and legislative preferences align, since both principal and agent favor simple authority structures. By contrast, when the legislative and executive disagree, outcomes are less predictable, and decision-making structures may be either simple or complex depending on the characteristics of the implementing agent (and in particular, whether the implementer is independent or subject to political control by the executive).

H4: The relationship between executive/legislative preference conflict and downstream allocation of authority should be attenuated when the policy proposal in question addresses complex or low-attention policy areas.

This idea - which represents an *interactive* relationship - also follows naturally from the theoretical foundations that underly **H1** and **H2**. Legislative and executive actors must both expend resources (cognitive and otherwise) in order to create complex authority structures. On high-salience policy issues in familiar substantive areas, both sets of actors will be highly incentivized to pay these costs. However, when addressing low-salience, complex policy problems, the costs incurred by creating a complex decision-making structure will likely overwhelm any payoffs that a higher-quality policy proposal may offer. As a result, in these situations the magnitude of the relationship hypothesized in **H3** will likely be small.

[2] Other studies (e.g. Bendor and Meirowitz 2004; Oosterwaal et al. 2012) have extended this idea to include expectations regarding *future* preference alignment. For simplicity, I omit considerations of this kind both in this section and in the empirical sections of this paper. However, incorporating expectations regarding future electoral performance represents a direction for future work.

3 A Relational Conception of Formal Power

3.1 From Text to Networks

In the social sciences more broadly, a common way to understand the contents of a particular corpus is to study the *relational* aspects of that document. 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)'s use a similar approach to analyze newspaper accounts of lynching episodes in the American South. Like GDELT, Franzosi et al. identify and code subject/action/object triples in their corpus, and record usage rates for various actors (African American victims; white citizens; law enforcement) and action types (violence; coercion; search; apprehension). This method reveals some surprising findings. For example, at least according to newspaper accounts, law enforcement officials were targeted for coercion by mobs nearly as frequently as African Americans, though the overwhelming majority of of violent actions was directed at lynching victims (Franzosi et al. 2012, 12).

This approach is similarly helpful in the legal setting. From a conceptualization standpoint, a natural way to understand the contents of a particular legal text is to examine the power relationships it contains. Focusing on these "implementing networks" reduces many substantively important questions to simple network-theoretic hypotheses. For example, work by Epstein and O'Halloran (1999), Volden (2002), Bendor and Meirowitz (2004), and others on relative allocation of authority to independent commissions and politically-controlled administrative agencies can be viewed as a question regarding the importance of each actor type to their respective implementation networks, operationalized through node centrality or influence metrics. Similarly, policy fragmentation (Kagan 2009; Farhang and Yaver 2016) can be operationalized using quantities like the number of nodes or average degree of the network. Better still, because these quantities can be calculated for any network, these kinds of quantities are straightforwardly comparable across any set of networks derived from some formal legal corpus of interest.

From a more practical standpoint, a relational conceptualization of formal power also helps to pinpoint the measurement challenges involved in a study 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 a natural measurement problem: in particular, how best should the relevant entities and edges be extracted from formal legislative documents? Many approaches to this problem are possible, but for the purposes of this dissertation I turn to a series of tools drawn from the natural language processing (NLP) literature. NLP is a catch-all term describing a set of computational methods that attempt to analyze the linguistic attributes of a given text. NLP methods are thus extremely wide-ranging, covering topics like part-of-speech tagging, grammatical parsing, lexical co-occurrence, latent content analysis (e.g. topic modeling), and much else besides. Coupled with a deep, politically- and legally-informed understanding of the documents in question, these tools can provide a powerful approach to the analysis of legal texts.

3.2 Entity Extraction

Entity extraction is a classic NLP problem, which has been attacked using a variety of heuristic and machine-learning approaches.[3] Here, I use a long short-term memory (LSTM) neural network, which is a common approach for many NLP problems.[4] Broadly, neural networks are a class of machine learning approaches which seek to predict some outcome of interest based on a series of "hidden layers", which iteratively manipulate some observed set of predictor variables in order to produce predictions. 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 and corresponding predicted outputs

[3] See, e.g., (see, e.g. Leetaru and Schrodt 2013) for an example of a specialized dictionary-based system, or the Stanford CoreNLP toolset Manning et al. (2014) for a general-purpose algorithm.

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

for nearby data points, creating a recursive, context-sensitive prediction structure ideal for analyzing sequentially-organized information.

From an implementation standpoint, I rely on the bidirectional long short-term memory (LSTM) neural network architecture proposed by Lample et al. (2016) and Ma and Hovy (2016).[5] Given a textual excerpt (e.g. a sentence or paragraph), this implementation predicts token-specific tags 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 surfaces relevant information for both simple and complex tagging rules. 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 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").

To generate training data for this model, I used a two-step procedure. First, I 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 supple-

[5] 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 blog post.

mented 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.

Second, I scraped the text of all enacted American legislation available through congress.gov, and used these legislative texts to construct a training set. For each text, I first segmented the text into sentences.[6] For each sentence, I 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.[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. This process left me with a training set consisting of some 29,080 sentences.

Using this set of training examples, I trained the LSTM model described previously, and assessed its performance.[8] 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,

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

[7]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.

[8]For details regarding hyperparameter specification, see 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.

predicted values for the held-out test set, and used these predictions to calculate predictive accuracy and F1 scores.[9] 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 to the one I employ, 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 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, since they approach the values offered in other work I suggest that they represent a strong starting point from which to work.

3.3 Relation Extraction

Compared with entity extraction, relation extraction is a substantially more complex 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. This additional information increases the scope of the prob-

[9] Two common metrics for assessing the performance of a classification algorithm are precision - defined as $P = \frac{TP}{TP+FP}$ - and recall - defined as $R = \frac{TP}{TP+FN}$ - where TP denotes the "true positive" detection rate, FP denotes the "false positive" rate, and FN denotes the "false negative" rate. The F_1 score is defined as $F_1 = 2\frac{1}{\frac{1}{P}+\frac{1}{R}}$, or the harmonic mean of precision and recall. This value is a common performance metric for imbalanced classification problems, since it rewards positive predictive accuracy in both the common and rare classes.

lem substantially, since it requires the model to consider larger blocks of text (e.g. words separating two named entities) and, for most applications, a larger typology of relationship types.

Fortunately, many variables of interest in the literature on formal institutional design do not rely on a detailed typology of relation types. For example, take (Farhang and Yaver 2016)'s work on fragmentation of authority. In their study, they define fragmentation as a tripartite concept, which counts the *number* of *distinct* actors empowered to execute a *particular statutory provision.* Though this definition might be enriched by a more nuanced definition of actor types or relationships between them, a more abstract notion of "fragmentation" still offers the authors rich theoretical ground with which to work. 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.*[10] 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.

The relation extraction procedure I employ in this paper, then, proceeds in two steps. First, I segment each text according to its internal organization (e.g. titles, sections, etc), and

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

remove preliminary material, section titles, and section headers from the text.[11] Second, I recombine each text into sections, extract entities from each section, and draw an edge between any set of entities that co-occur in a given section. These two components combine to form the extracted "implementing network" for each law, which forms the basis for the other analyses I present.

4 The Design of American Law

4.1 Constructing Implementing Networks

As I describe at the outset of this paper, my primary goal in this project is to study allocation of authority patterns in enacted American legislation. I therefore constructed an original dataset consisting of all enacted national-level legislation texts and metadata available through congress.gov, the official U.S. Congress legislative database.[12] Temporally, this dataset approximately covers the period from 1990-2014.[13] Using the Congressional Bills Project's metadata set, I then filtered so-called "commemorative" bills from the dataset, leaving me with a total dataset of 3467 observations (out of 4800 total).[14]

For each text, I first stripped administrative information (e.g. date of passage; legislative history; transcription notes), and segmented each document into sections.[15] I then removed the first section from each document. In contemporary American legislation, the first section of each document always contains a set of preliminary material, such as an official "short title", a table of contents, or similar information, which is not relevant for the analytical task I undertake in this section. Next, I extracted entities from each document using the pre-

[11] Using the parser contained in https://github.com/rbshaffer/constitute_tools.

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

[13] 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.

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

[15] 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.

trained LSTM model I describe in §3, and discarded all named entities that were mentioned only once in their respective documents.[16] Finally, I recombined 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 aid intuition, I provide two example outputs from this process in Figures 1 and 2. Beginning with the simpler case, the Enhanced Partnership with Pakistan Act of 2009[17] is a relatively straightforward foreign aid bill intended to provide military and developmental assistance to the Government of Pakistan. The law authorized the 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.[18] 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)[19] 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 visual-

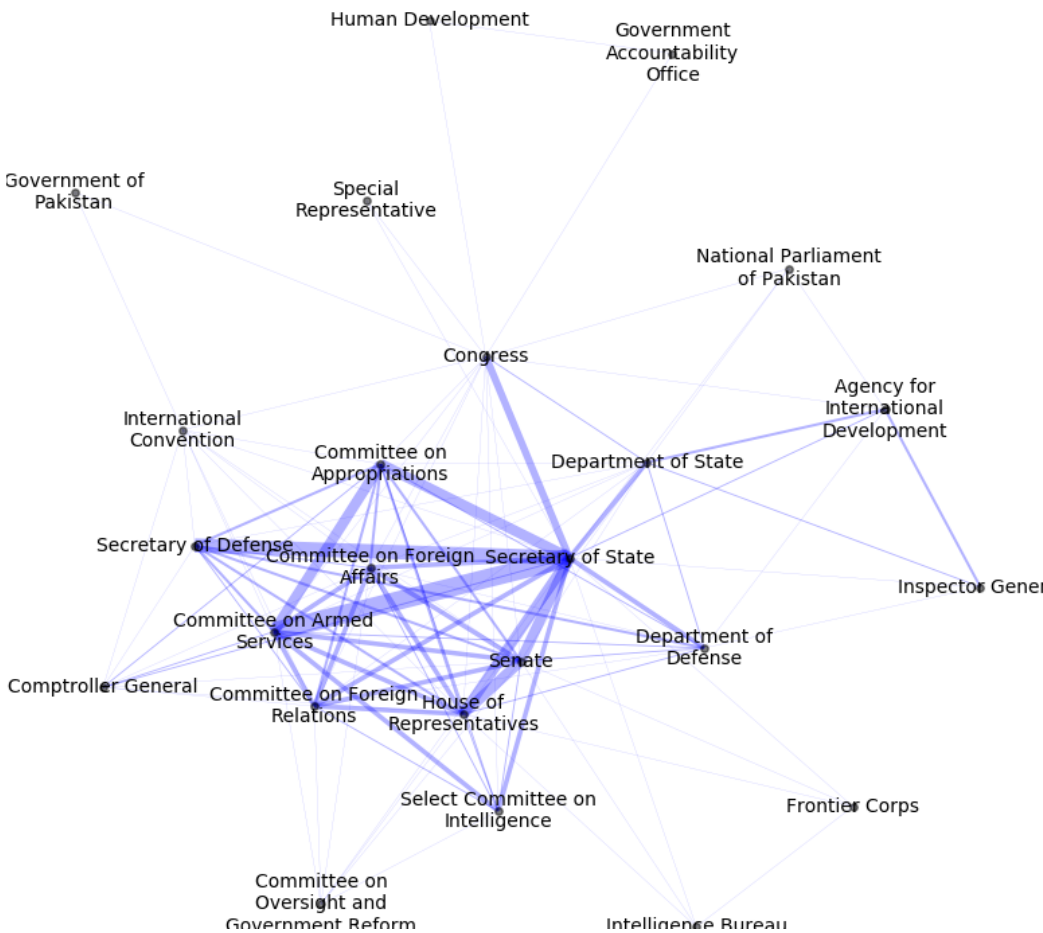
[16] This heuristic is drawn from the natural language processing literature (see, e.g., Grimmer and Stewart 2013), and is useful in cases where typographic errors or other types of false positives are likely to be common.

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

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

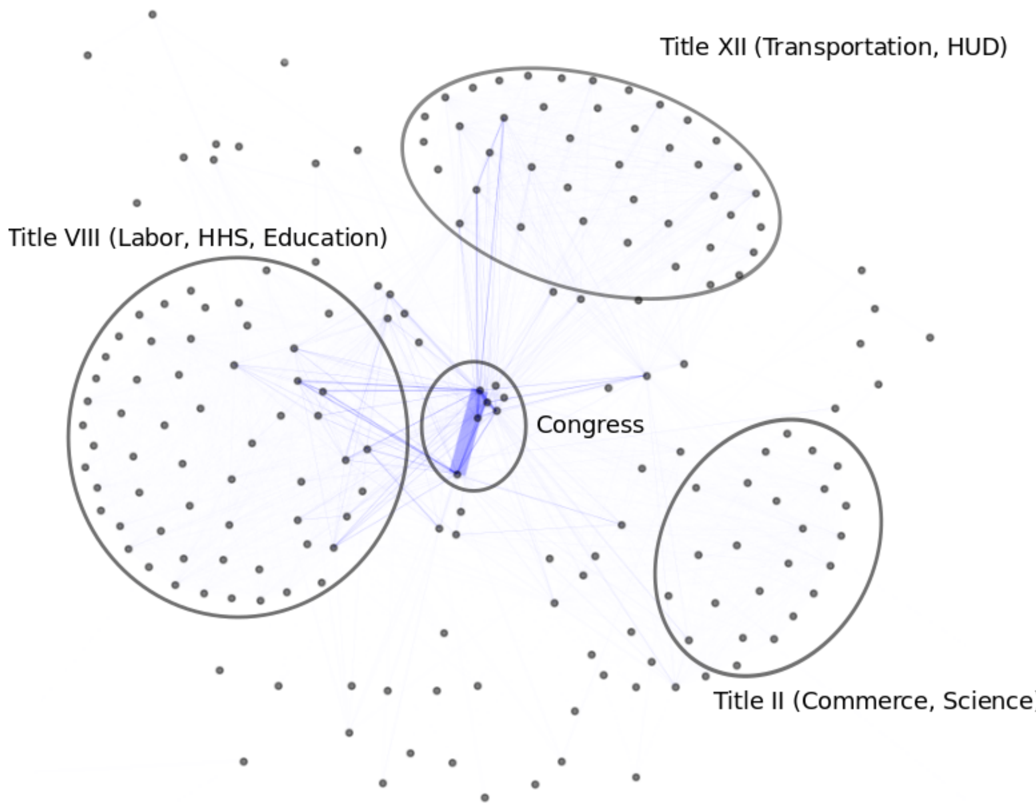
[19] Public Law No. 111-5. For original text see corresponding congress.gov page.

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.

ization 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 organized 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.

4.2 The Dependent Variable: Network Fragmentation

Despite its theoretical importance, relatively few authors have operationalized fragmentation in a systematic fashion. McCubbins (1985) and Kagan (2009), for example, offer broad overviews of bureaucratic organization and types of constraints on implementor discretion, but little in the way of empirical operationalization of fragmentation or related ideas. Farhang and Yaver (2016) offer perhaps the most systematic description of the concept, defining "fragmentation" 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.

Fragmentation, from this perspective, therefore refers to the *number of actors involved in the execution of any particular statutory function.* Farhang and Yaver (2016) operationalize this idea a fairly direct fashion. For each statute they examine, they read the statutory text and code (1) the number of actors empowered to execute a core regulatory function in the law, (2) the number of federal agencies empowered to execute a core regulatory function in the law, and (3) the number of instances in which two or more actors or agencies were given simultaneous authority over a particular policy area. They argue that these ideas all tap aspects of their underlying concept of interest (which they further support using visual and statistical evidence), and therefore use a simple average of these three components as the final dependent variable in their analyses.

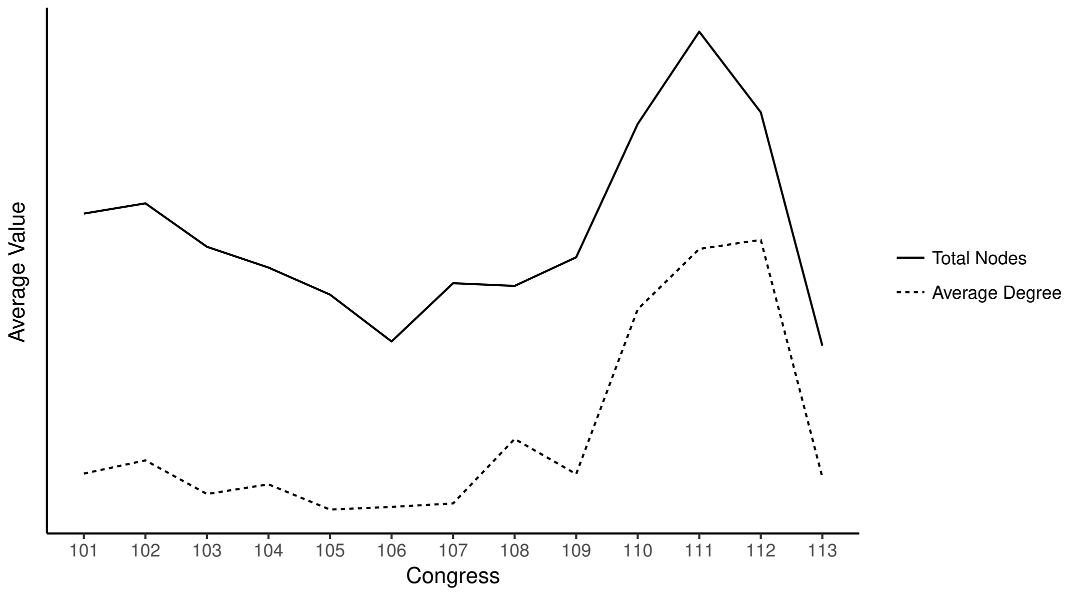
Translated to the network context, this conceptualization scheme offers a natural starting place from which to operationalize fragmentation of implementing authority. As Farhang and Yaver (2016) note, both a simple count of the number of actors/federal agencies tasked with implementing key policy programs and the number of instances in which actors are offered overlapping policy authority within a particular statute offer reasonable proxies for the extent to which authority within that statute is fragmented.

Fortunately, both of these ideas are operationalizable using the tools I develop in §3. Since entity extraction is a substantial part of the network-based measurement scheme I propose, a simple way to count the number of actors involved in the execution of each statute is to count the *number of unique entities* identified in each law. Though this measure does not focus on actors involved in "core regulatory functions" as in Farhang and Yaver (2016)'s measurement scheme, it still likely taps the same underlying concept as the more focused count these authors employ. Measuring the frequency with which multiple actors are empowered to execute the same statutory function is more complicated, but a straightforward approach is to calculate the *average degree* of a bill's implementation network, defined as the average edge weight for each node.[20] Larger values on both statistics indicate a more fragmented implementing network, while smaller values indicate a more "siloed" implementing structure.

Like Farhang and Yaver (2016) I find that both of my candidate measures 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 substantial overdispersion (maximum values of

[20]With the average degree defined to be zero for networks with zero named entities.

Figure 3: Comparison of two measures of legislative fragmentation.



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[21] leads to a pairwise correlation value of 0.95. Because of this overlap, for the remainder of this paper I focus on results generated using the node-count measure of my dependent variable, since this variable is simpler to model.

4.3 Predictor Variables

As I describe in §2, my key theoretical contention is that previously observed relationships between partisanship and downstream design of legislation should be moderated by policy area and by the public salience of the bill in question. 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 and legislative branches at the time that the bill was passed. Based on existing theory, on average we should expect bills passed under unified government should contain less fragmented im-

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

plementing structures than their counterparts passed under divided government.

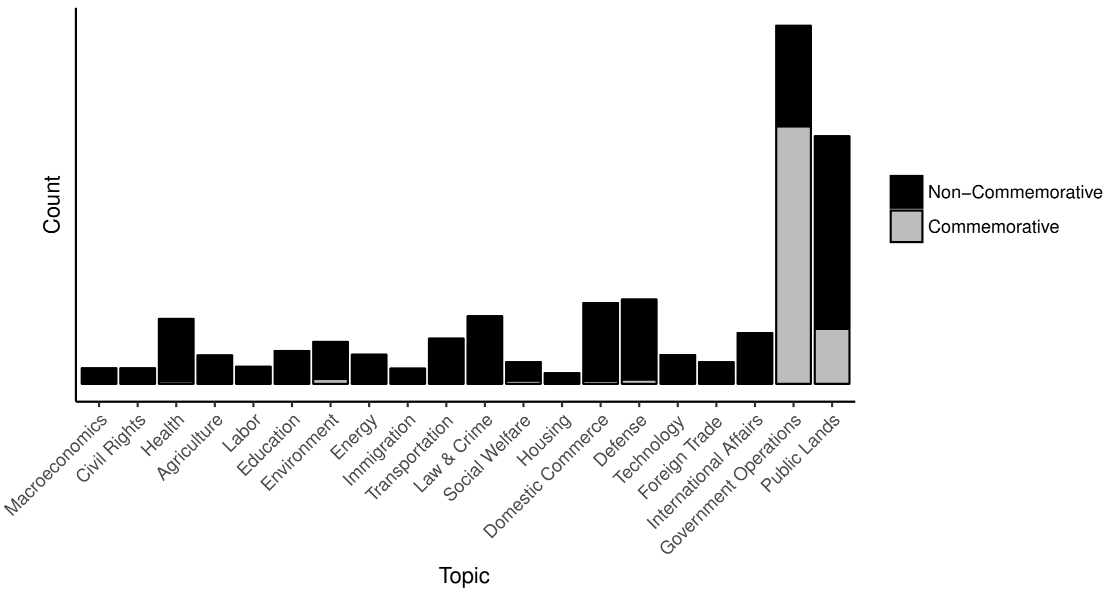
Besides this basic partisanship variable, I also include two individual-level and two bill-level covariates. For my individual-level variables, I include covariates corresponding to the DW-NOMINATE score of the proposing member and an indicator variable denoting whether the proposing member was a part 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 my 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 variables. First, to measure the broader public salience 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 authors of these bills may not have the time to gather a substantial number of cosponsors.[22] As an additional

[22]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.

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.

control, I therefore include a dummy variable indicating whether the bill in question was an annual appropriations bill or a defense authorization bill, which should serve 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 *negative* on average, though this relationship may vary some by policy area.

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

5 Modeling

To model bill-level fragmentation patterns, I employed a Bayesian hierarchical negative binomial model, with a hurdle component. 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.[23] Within each part of the model, I estimate coefficients corresponding to each of the six variables I identify in Table 2.[24] 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 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[25] or bills which consist of technical amend-

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

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

[25]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.

ments, corrections, or updates to other pieces of legislation.[26] 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 hurdlecomponent of the model serves to separate 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 hierarchical modeling structure allows for this kind of variability, offering additional analytical leverage.

[26]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.

6 Results

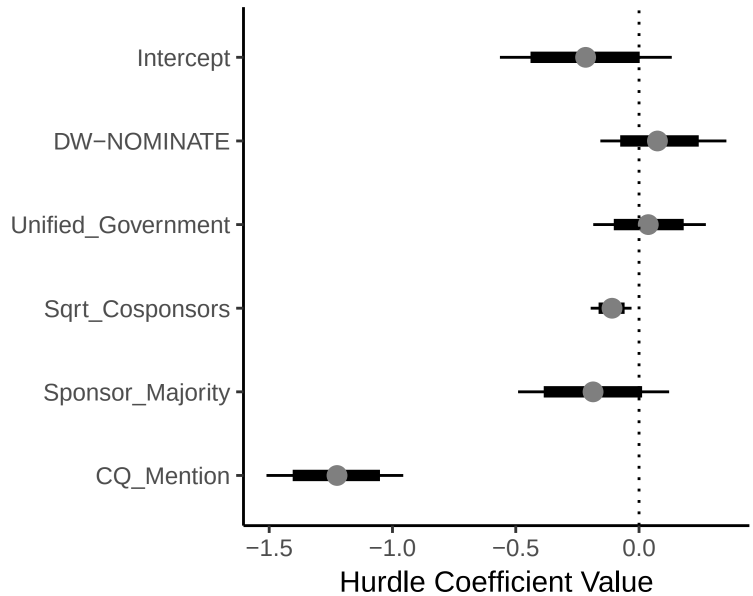
6.1 Hurdle Model

Top-level posterior means and credible intervals for the hurdle model coefficients are given in Figure 5. 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, 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

Figure 5: Top-level coefficients for the hurdle model.



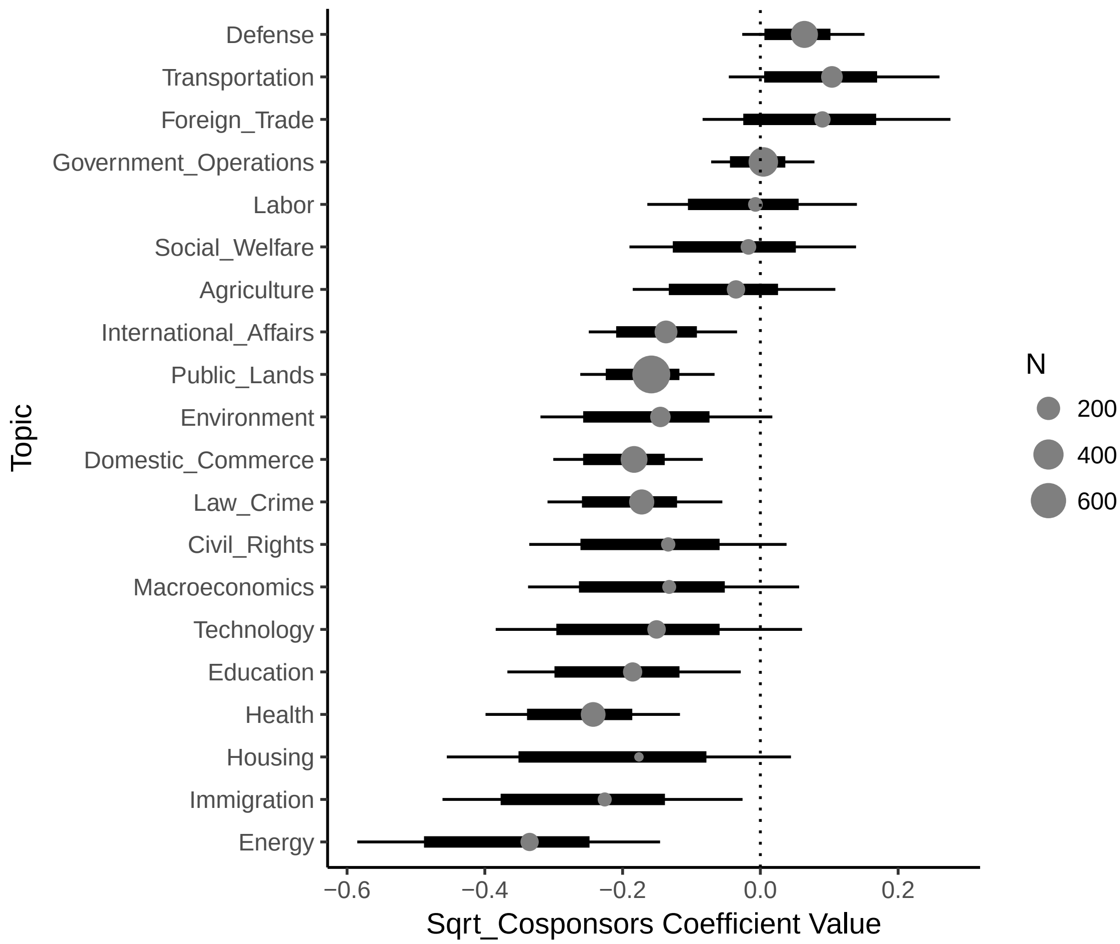
Dots indicate posterior mean values. Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Positive estimates indicate that an increase in the given parameter value makes the dependent variable more likely to take on a value of zero.

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 fragmented 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.

6.2 Count Model

Broadly, the count component of the model I present - which is restricted to bills that actually affect administrative authority - can be interpreted similarly to the hurdle component. Since the count component uses a log-link, we can interpret the exponentiated coefficient

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



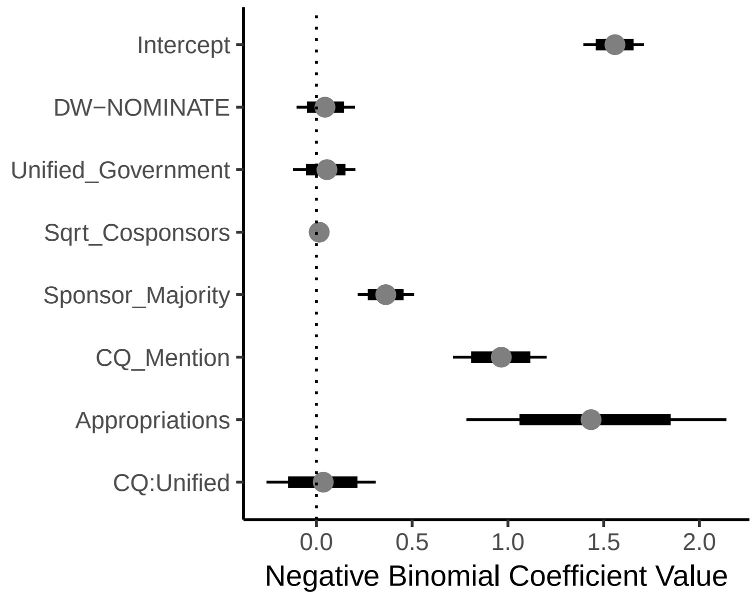
Dots indicate posterior mean values. Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Negative estimates indicate that an increase in the given parameter value makes the dependent variable less likely to take on a value of zero. Dot sizes scaled by the number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

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 fragmentation when comparing majority-sponsored bills to their minority-sponsored counterparts. Using a similar procedure for the `Appropriations` variable yields a predicted $\approx 320\%$ increase in fragmentation. 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 fragmented.

The coefficient on the `CQ_Mention` and `Unified_Government` variables partially match my expectations, but contain some unexpected results. As predicted, publicly salient bills are substantially more fragmented than their non-salient counterparts. Contrary to my expectations, however, this relationship is essentially identical under divided and unified government. Averaged across policy areas, a mention in CQ's year-end almanac corresponds to a $\approx 163\%$ increase under divided government and a $\approx 170\%$ increase under unified government. Most surprisingly of all, bills passed under divided and unified government display essentially no differences in fragmentation levels.

As in the hurdle model, focusing on top-level coefficients can conceal substantial effect heterogeneity. For example, consider the `DW-NOMINATE` variable. Averaged across policy areas the ideological orientation of the proposing member has little impact on the a bill's administrative structure. However, this broad view conceals some potentially interesting policy area-specific differences. Since the `DW-NOMINATE` variable is scaled from -1 (most liberal) to 1 (most conservative), positive coefficients indicate that more conservative members tend to propose more fragmented bills in a given policy area, while negative coefficient estimates indicate that more liberal members tend to propose more fragmented bills in a given area. As Figure 8 shows, most policy areas display coefficient values near zero on this scale. However, bills addressing defense policy issues tend to show positive coefficient values, while bills addressing technology, and law/crime display negative estimates. Though these estimates

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



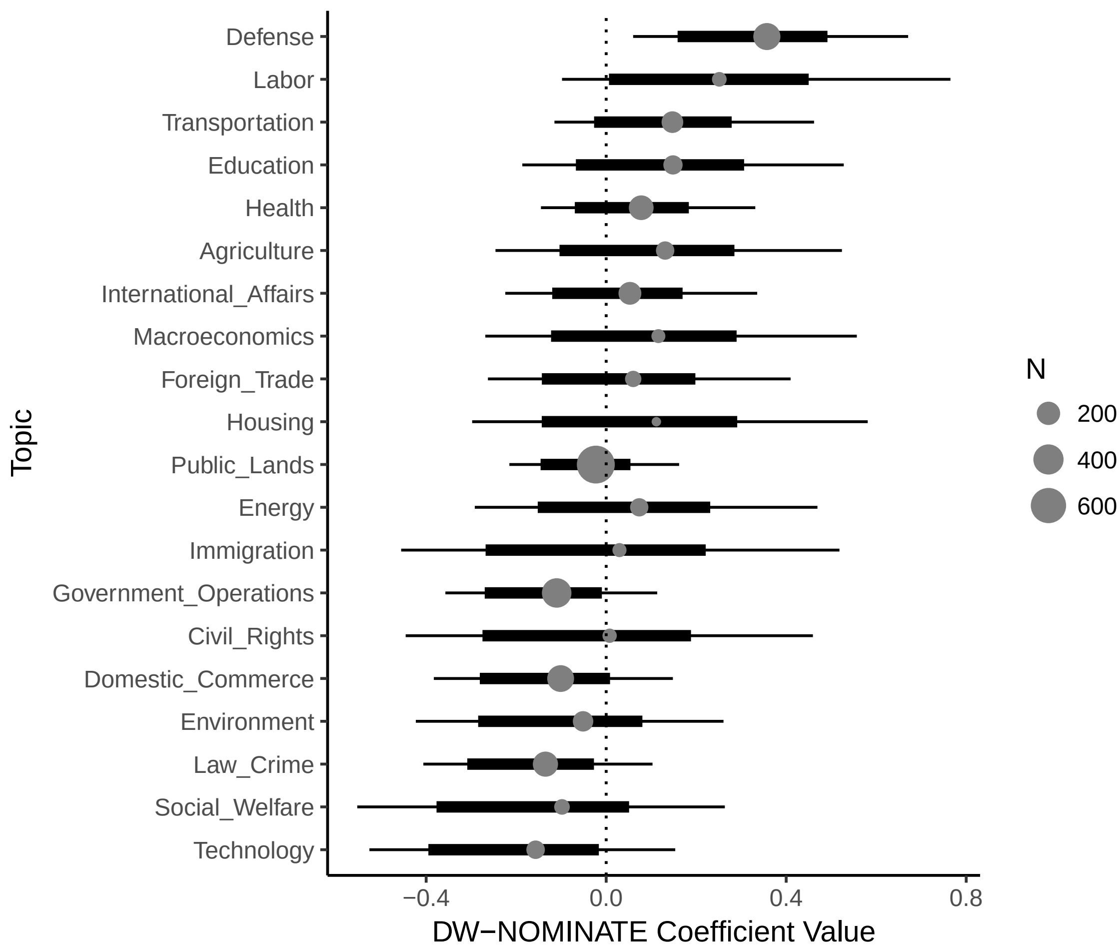
Dots indicate posterior mean values. Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Positive estimates indicate that an increase in a given coefficient increases fragmentation.

are high-variance, their effect sizes are substantively significant; for example, a one-standard deviation increase in the `DW-NOMINATE` variable (≈ 0.45) would be predicted to produce a $\approx 18\%$ increase in fragmentation for a defense bill, and a $\approx 15\%$ decrease in fragmentation for a technology bill.

This pattern roughly tracks the interaction between ideology and policy priorities I predict in §2. Designing complex implementing structures is costly for lawmakers in both cognitive and political terms. As a result, lawmakers are only likely to pay these costs when they are appropriately incentivized. Intuitively, we should expect conservative lawmakers to be differentially motivated to create more complex institutional structures on bills addressing conservative policy priorities (e.g. defense) and liberal lawmakers to be similarly motivated in liberal policy priorities (e.g. social welfare, environment). Because most coefficient estimates on this variable possess a low level of precision, these conclusions are tentative; however, I argue that these results offer at least suggestive support for the hypotheses I advance.

This pattern of strong effect heterogeneity also helps explain some of the surprising top-

Figure 8: Second-level DW-NOMINATE coefficients, count model.



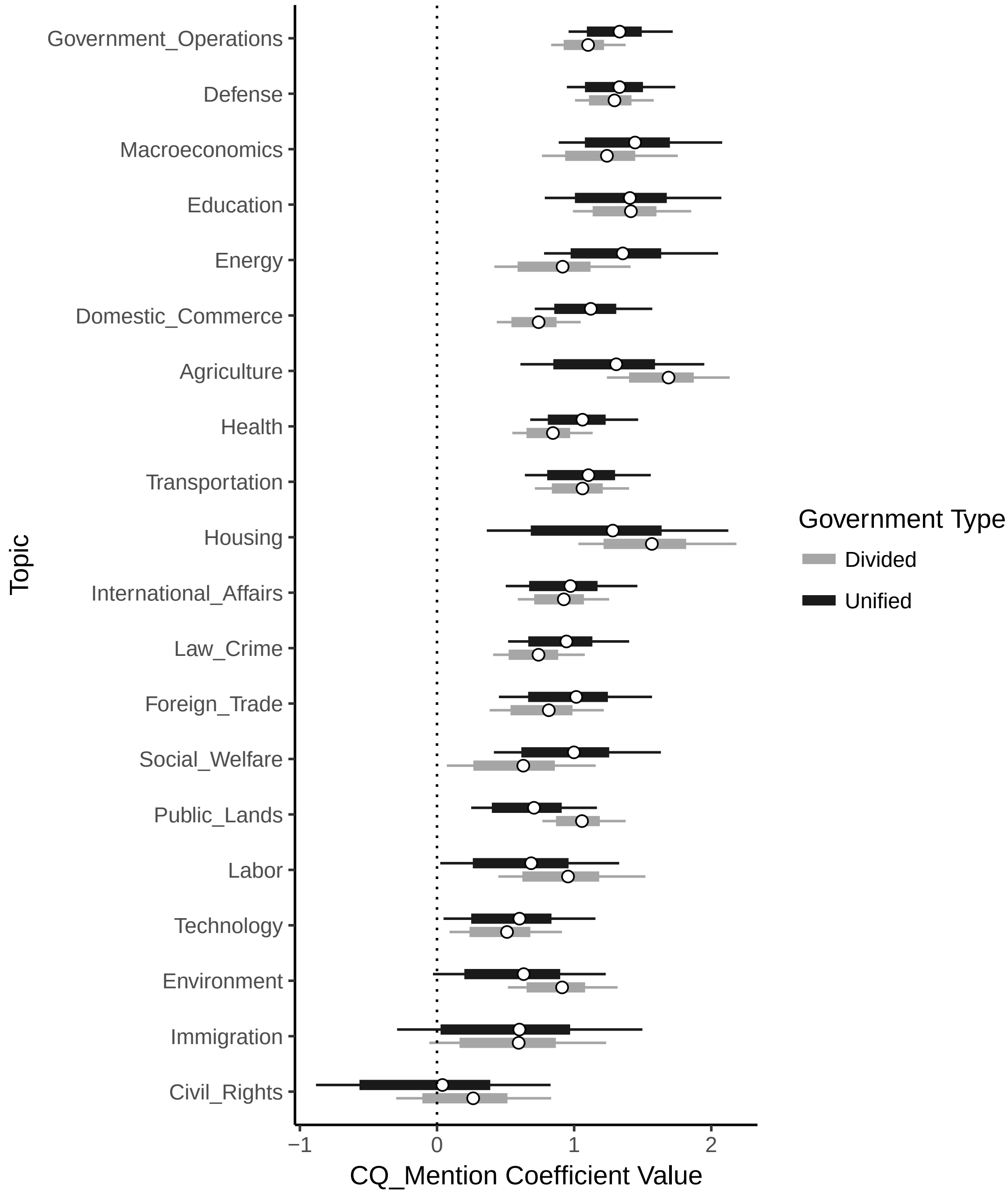
Dots indicate posterior mean values. Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Positive estimates indicate that an increase in the coefficient increases the bill's estimated fragmentation value. Dot sizes scaled by the number of bills in each policy area.

level results in the CQ_Mention and Unified_Government variables. Since I interact these two variables, 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 simply 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 9. As shown in Figure 9, though this variable's estimated effect is large and positive in for most policy areas, the scale of its effect varies dramatically. 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 the average effect size is larger and rankings across policy areas are somewhat shifted.

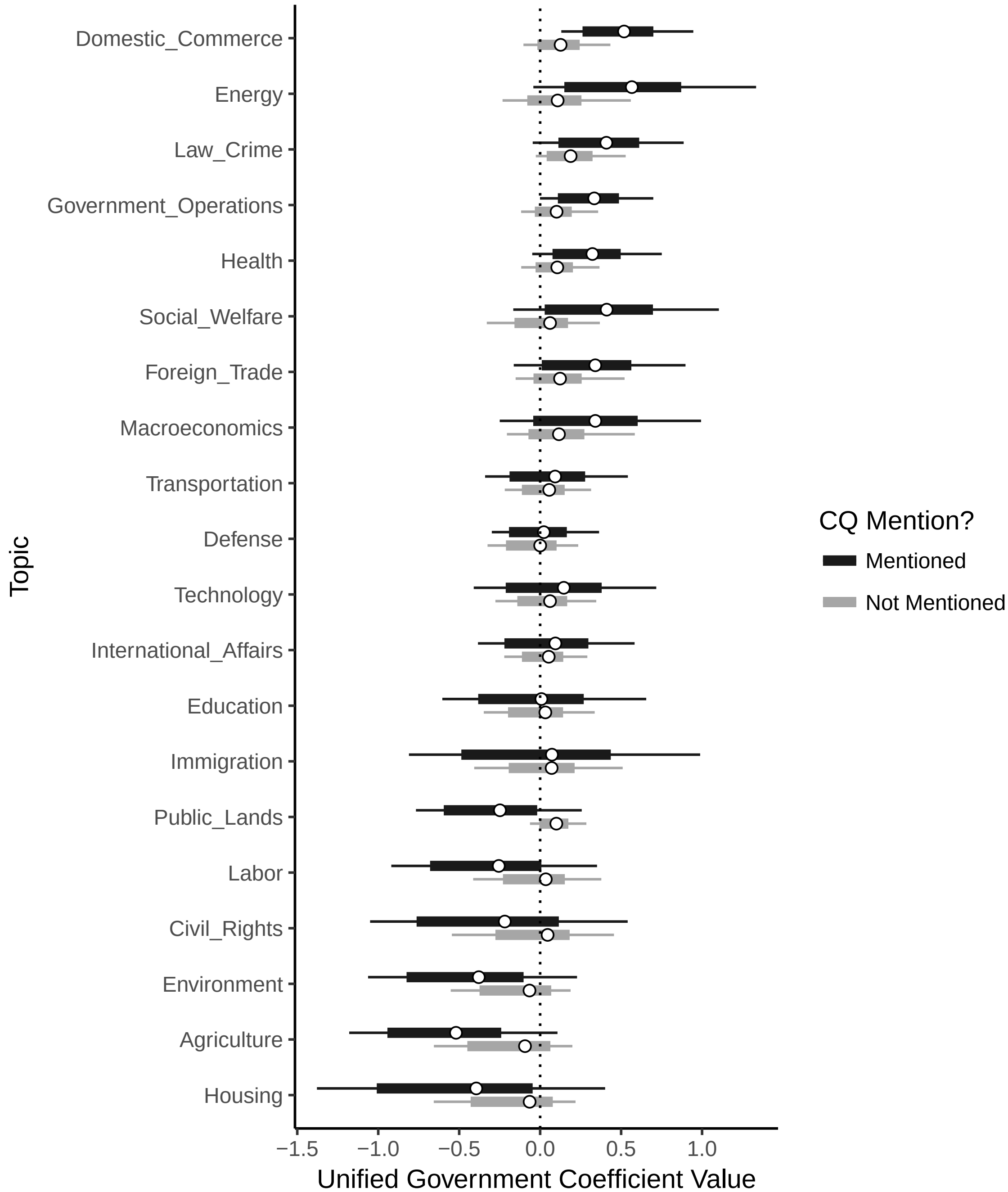
As shown in Figure 10, we can use a similar strategy to investigate effect heterogeneity in the Unified_ Government variable. As I predict in §2, for non-salient bills the relationship between partisanship and downstream design of legislation is attenuated. In other words, bills that receive little or no public attention possess essentially equivalent implementing structures when passed under unified or divided government. By contrast, salient bills passed under unified government are somewhat less fragmented than their counterparts passed under unified government; however, this effect is limited to approximately a third of the policy areas I identify. In the remaining two-thirds of policy areas, the predicted effect of moving from divided to unified government is either zero or actually positive. This effect is largest for domestic commerce and social welfare bills, but is also noticeable for health bills.

Figure 9: Second-level `CQ_mention` coefficients, count model. Estimates for both unified and divided government scenarios included.



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.

Figure 10: Lower-level `Unified_Government` coefficients, count model.



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.

These findings help explain the surprising top-level results I present in Figure 7, and offer partial support for the hypotheses I offer at the outset of this section. However, the overall weakness of the relationship between divided and unified government and downstream allocation of authority is surprising, and merits further investigation. Interestingly, the results I present here actually align with those from at least one other major study. In their book, Huber and Shipan (2002) limit their attention to state-level health policy bills, and find a positive relationship between discretion (an idea related to fragmentation) and unified government. This finding matches the coefficient I present for health bills, and offers some reassurance that coefficients I report are accurate. One possible explanation for the difference between bills addressing the policy areas with positive coefficients on the `Unified_Government` variable and the remainder of the dataset relates to the type of bills passed in these areas. Since bills in health, transportation, government appropriations, and domestic commerce all frequently involve substantial appropriations, lawmakers acting under unified government may be more willing to fragment authority on these bills in order to protect their favored policy programs. Under divided government, by contrast, appropriations bills may create fewer new programs, creating fewer incentives for lawmakers to create complex implementing structures for these bills. Clearly, however, probing this and other possible explanations further represents a direction for future research.

Two other possible explanations for this divergent finding are *time* and *selection*. Since Epstein and O'Halloran (1999) and Farhang and Yaver (2016) both study legislation over the period from 1945 onwards, the strong divided government effect they identify may actually be restricted to the earlier period in their dataset. However, this explanation seems unlikely. If a divided/unified government effect is present, increasing Congressional polarization in recent decades should *increase* rather than *decrease* the size of this coefficient, leading me to observe a larger coefficient on this variable than that observed in other studies.

A more likely explanation for the divergence between my findings and those of most previous studies is a *selection* problem. Because Epstein and O'Halloran (1999) and Farhang and Yaver (2016) only examine laws drawn from Mayhew (1991)'s list of "significant" legislation, their studies necessarily select out all but the highest-attention bills in the dataset. However, based on the expectations I present in §2, this set of bills is precisely the set in which we should expect to see the strongest relationship between divided government and fragmentation of authority. When addressing high-salience policy areas, politicians are particularly incentivized to design administrative structures carefully and to be suspicious of executive malfeasance. However, since bills that reach this level of public salience are relatively rare, in most cases considerations like policy area and partisan policy priorities are more influential on the implementing structure in a given bill. Re-estimating the model I present to only include high-salience bills might help to probe this explanation. However, a narrower measure of public salience than the one I employ might be necessary in order to observe this effect.

7 Conclusion

Overall, the results I present provide support for many of the hypotheses I present in §2. Throughout the models I present, the *salience* of a given bill (as measured by mention in CQ's year-end almanac and the number of cosponsors that bill attracts) consistently represent some of the most important predictors in the model. High-salience bills are dramatically more likely both to affect administrative jurisdiction and to contain more fragmented implementing structures.

Also as predicted, many of the relationships I examine are conditional on *policy area*. In the hurdle component of the model, for example, increasing cosponsor counts decrease the probability that a bill will be in the "non-administrative" category for most policy areas, but actually increase this probability for defense bills and other bills in policy areas with many measures passed outside the usual legislative process. In the count component of the model, the unified government coefficient is also contingent on policy area. Averaged across policy areas, unified-government bills are no more or less fragmented their counterparts passed under divided government. But, for politically salient bills, the scale and direction of this relationship actually varies substantially across policy areas, with bills in some policy areas displaying a negative relationship between unified government status and fragmentation and some displaying a positive relationship. Probing this finding represents a direction for future research.

These results emphasize the importance of both the theoretical ideas and the measurement techniques I introduce in this dissertation. 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 develop in this dissertation, these findings would not have been possible to produce, emphasizing the importance of scalable, broadly applicable measurement techniques for applied work.

The work I present here leaves a number of directions for future research. Within the domain of American legislation, comparing enacted legislation to proposed but unpassed bills would allow me to compare majority- and minority-proposed legislation in a more direct fashion, without selecting for passage. Similarly, comparing the changes in enacted bills across drafts would offer additional insights into the negotiating process underlying each bill, and would allow me to identify the contributions of each individual legislator in a more direct fashion. Last - but certainly not least - the approach I develop also offers opportunities to study lawmaking patterns under different systems of government, allowing me to expand the scope of this project beyond the American context.

References

Baumgartner, F. R., Breunig, C., Green-Pedersen, C., Jones, B. D., Mortensen, P. B., Nuytemans, M., and Walgrave, S. (2009). Punctuated equilibrium in comparative perspective. *American Journal of Political Science*, 53(3):603–620.

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*.

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. (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.

Gailmard, S. and Patty, J. W. (2012). *Learning while governing: expertise and accountability in the executive branch*. University of Chicago Press, Chicago.

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.

Kagan, R. A. (2009). *Adversarial legalism: The American way of law*. Harvard University Press.

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. (1985). The legislative design of regulatory structure. *American Journal of Political Science*, pages 721–748.

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.

Oosterwaal, A., Payne, D., and Torenvlied, R. (2012). The effect of political disagreement on discretion mechanisms and conditions. *Administration & Society*, 44(7):800–824.

Paas, F., Renkl, A., and Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational psychologist*, 38(1):1–4.

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*.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4):295–312.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review*, 22(2):123–138.

Volden, C. (2002). A formal model of the politics of delegation in a separation of powers system. *American Journal of Political Science*, pages 111–133.

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

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

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.[27] 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.

[27] 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.[28] 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

[28] 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.[29]

Following Gelman et al. (2014), in Figure 11 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.

[29]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 11: Posterior predictive plots for the node count dependent variable