

Using Ideas As My Maps: Tracing the Flow of Ideas Through the Legislative Process

by

Joshua Y. Lerner

Department of Political Science
Duke University

Date: _____

Approved:

Mathew D. McCubbins, Co-Supervisor

John H. Aldrich, Co-Supervisor

David W. Rohde

Michael C. Munger

Justin R. Grimmer

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor
of Philosophy in the Department of Political Science in the Graduate School of Duke
University 2017

Abstract

Using Ideas As My Maps: Tracing the Flow of Ideas Through the Legislative Process

by

Joshua Y. Lerner

Department of Political Science
Duke University

Date: _____

Approved:

Mathew D. McCubbins, Co-Supervisor

John H. Aldrich, Co-Supervisor

David W. Rohde

Michael C. Munger

Justin R. Grimmer

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Political Science in the Graduate School of Duke University 2017

Copyright © 2017 by Joshua Y. Lerner
All rights reserved except the rights granted by the
Creative Commons Attribution-Noncommercial Licence

Abstract

Ideas are an often-ignored part of the political science literature on legislation. Most studies of Congress focus on institutional concerns, party considerations, gains from exchange, and other unideated concepts. But the exchange of ideas in the writing of bills represents the fundamental realization of public policy in practice, and any story of major legislation that does not adequately deal with where the ideas in said statutes come from is missing a vital part of the story. Particularly, the role in Congressional hearings as an information exchange is often given only prima facie concerns in legislative expertise models, but what exactly is being transmitted is almost never addressed directly.

In the first section, I study how think tanks position themselves strategically to influence Congressional behavior. Using a dataset of think tank citations from Congressional floor speeches and committee testimony records, I compare the influence of think tanks based on a new measure their ideology, and in doing so, show that think tanks engage in strategic ideological positioning to maximize their influence. An additional hypothesis examined is the relationship between think tank members' previous work experience in government with the organizations' overall prominence. By treating think tanks as strategic actors in the complex constellation of legislative politics, this chapter makes the argument that think tanks need to be considered in even the most cursory account of the policy making process.

In the second section, I introduce an approach to idea tracing that is better able to answer the questions: whom does Congress listen to and when during the legislative process do they listen? Using machine learning and text-as-data approaches, I establish a more

direct framework to evaluate the impact of testimony on the language of legislation through the development a text reuse approach to map specific ideas from congressional hearings to bills. Specifically, I focus on the language of Dodd-Frank and systematically trace the development of ideas with calls to testify in Congress from experts. In doing so, I provide the first account of who is being listened to during the legislative process.

The final substantive section expands on the previous section and applies the technique developed more broadly, to multiple pieces of important financial legislation. Using a topic model to identify major policy areas, I establish a more direct causal framework to evaluate the impact of testimony on the language of legislation. Utilizing a modified text-reuse method which combines a localized alignment text-reuse algorithm for identifying reused ideas, I compare the rate of idea reuse between hearings and final bill language for four major bills, spanning four Congresses. This allows for an effective way to test which speakers are being listened to during hearings, and under what conditions expert testimony has an impact on bill writing. By focusing on financial legislation, this paper can directly compare the work of people in multiple Congresses and, by discussing multiple bills, see when these speakers matter and when they don't, and in doing so, model a network of influence of external speakers in Congress.

To my mother, whose guidance got me here in one piece;
To my father, of blessed memory, who believed in me beyond all else;
And to Catherine, always and forever.

Contents

Abstract	iv
List of Tables	viii
List of Figures	ix
List of Abbreviations and Symbols	xi
Acknowledgements	xii
1 Introduction	1
2 Getting the Message Across: Evaluating Think Tank Influence in Congress	4
3 Using Ideas as My Maps: Exploring the Flow of Ideas From Hearings to Legislation	30
4 Money Doesn't Talk (It Swears): Tracing Ideation in Major Financial Legislation	56
5 Conclusion	83
A Extras to Chapter 2	85
A.1 Appendix: More information about Board of Directors of Think Tanks . . .	85
A.2 Appendix: Different Count Regression Methods	87
B Extras to Chapter 3	90
C Extras to Chapter 4	99
Bibliography	106
Biography	118

List of Tables

2.1	Estimating Citations and Testimony in Congress	23
3.1	An Example of Text-Reuse Between Testimony and Bill Language	41
3.2	Dodd-Frank Testimony Totals by Speaker Organization	44
3.3	Dodd-Frank Section Totals by Version and Date	45
3.4	Text Reuse Modeling	50
4.1	Major Financial Legislation for chapter	73
4.2	Structural Topic Model on Hearings and Bill Sections Policy Topics	76
4.3	Major Financial Legislation and Reuse Rates	77
4.4	Modeling Text Reuse of Financial Legislation with Select Topics	79
A.1	Distribution of CFScore for Think Tanks	86
A.2	Alternative Citations Models	87
A.3	Alternative Testimony Models	88
A.4	Citations by Year per Think Tank	89
C.1	Structural Topic Model on Hearings and Bill Sections w/74 Topics	100

List of Figures

2.1	Comparison between Groseclose-Milyo Scores and CFscores for Think Tanks	17
2.2	Distribution of Congressmen and Think Tanks by CFscore	18
2.3	Estimated Citations and Testimony holding expenditures, age, and connected employees at the median, with no lobbyists.	24
2.4	Estimated Citations and Testimony holding expenditures, age and connected employees at the median, with no lobbyists.	25
3.1	Illustration of the Structural Topic Model (Roberts, Stewart, Airoldi 2014) .	39
3.2	Topics for Combined STM	48
3.3	Predicted Counts by Topic group: Low Complexity Topics	51
3.4	Predicted Counts by Topic group: High Complexity Topics	52
3.5	Predicted Counts by Principal Component of Topics	53
4.1	Predicted Counts by Oversight Topic over Bill	80
4.2	Predicted Counts by Regulatory Topic over Bill	81
B.1	Topic Model for Testimony Only	91
B.2	Changes to Federal Reserve Topic	92
B.3	Changes to Regulation and Rules Topic	93
B.4	Topic Model for Bill Only	94
B.5	Changes to Amending Topic	95
B.6	Changes to Federal Banking Topic	96
B.7	Alternative Specification of Topics: Invest Topic	97

B.8	Alternative Specification of Topics: Federal Reserve Topic	98
C.1	Evaluation of K 's for STM part 1	101
C.2	Evaluation of K 's for STM part 2	102
C.3	Banking Topic Over All Bills	103
C.4	Delegation Topic Over All Bills	103
C.5	LASSO of Topics on Text Reuse	104
C.6	LASSO of Topics on Text Counts	105

List of Abbreviations and Symbols

Abbreviations

GLBA	Gramm-Leech-Bliley Act
SOX	Sarbanes Oxley Act
TARP	Troubled Asset Relief Program
DF	Dodd-Frank Act
TT	Think Tank
CFscore	Contributions based Ideal Point (from Bonica 2014)
LDA	Latent Dirichlet Allocation—Topic model
STM	Structural Topic Model
MC	Member of Congress

Acknowledgements

This work could not have been done without the direct and oftentimes forceful vision of my advisor, Professor Mathew McCubbins. His supervision and guidance was invaluable. I would also like to thank the rest of my committee: Professors John Aldrich, David Rohde, Michael Munger, and Justin Grimmer, all of whom have shared with me tremendously thoughtful advice, friendly guidance, and general attention that I could never repay. I could not be the person or scholar I am today without all of their attention.

I would also like to thank my wife who has made innumerable sacrifices over the years while I pursued this, sometimes quixotic, goal, and only rarely found fault.

Ah, but I was so much older then
I'm younger than that now

Bob Dylan

Introduction

Ideas, which have played such an important role in historical/philosophical writings on politics, are mostly ignored in the modern social sciences. Instead, an emphasis is placed on more concerns that can be modeled more directly and tangibly. A return to a literature of “ideas” in the social sciences does not, however, mandate a return to non-empirical or normative studies, but, as I argue, can be integrated into a more positivistic, scientific paradigm. Developments in the computational study of language have allowed for more objective approaches to modeling how people think, speak, and write about topics, and in doing so have unlocked a new, facile way to address how we communicate and what is being communicated (Blei 2012).

In my dissertation, I will take these computational approaches to the study of language and attempt to trace the flow of ideas in legislation. In doing so, I introduce a novel way of approaching both legislative politics specifically, and politics and policy generally. This project addresses the fundamental ways in which legislation is shaped by ideas and brings with it the prospect of a major rethinking of how our laws come to be.

This project addresses one of the fundamental questions to the functioning of democratic governments: who is being listened to when Congress writes laws? While this project does

not deal with the voter side of this arrangement, it does address the content of legislation being passed, and, in doing so, speaks to the basic ways in which the legislature implements public policy concerns. The creation and diffusion of policy ideas in the legislative process is fundamental to understanding exactly the ways in which the legislature is writing and implementing the laws that are part-and-parcel to the successful implementation of democratic will. If we care at all about how legislators accurately create policies that both represent the public at large and do their best to improve the general welfare of the nation, then we have to care about the substance of the bills being passed and the idea-generating process behind them.

This dissertation is brought to bear by three different motivating questions: Whom is Congress listening to? When are they listening? And to what end? Each of these chapters deals with one or more of these questions and explores how external forces shape the idea environment from which Congress legislates.

Chapter 2 addresses the first questions by looking at think tank positioning in Congress. Whose experts get called upon to testify before Congress, and which organizations do members of Congress cite when trying to justify whatever legislation they are writing or attempting to pass? Which sets of experts, as the title says, are able to “get the message across” by winning a seat at the table? This section studies this from the think tank perspective but views the interactions between think tanks and members of Congress as a strategic game in which think tanks position themselves to maximize their ideological proximity to the members of Congress, all while not sacrificing the appearance of objectivity in their writings. This chapter sets the precedent upon which the next two chapters respond, namely once we have *who is speaking*, next we need to know *are they being listened to?*

Chapters 3 and 4 are both designed to answer the question: who is being listened to and when? Pursuant to these questions, Chapter 3 does three novel things. First, in it, I develop an approach to tracing ideas in legislation based on direct contextualized text similarities, taking advantage of modern computational capacity so it is conceptually straightforward and

largely portable. The second thing that this chapter accomplishes is to build and develop a general theory of idea transmission that seriously treats ideas as a type of information, which allows us to think of it in computationally. This theory also builds off of Congressional theories of committee participation and motivations. The final novel element of Chapter 3 is the evidence as to whom members of Congress listen to when writing legislation (somewhat beyond the conventional story normally told). Chapter 3 is really an introduction to a general theory and approach to idea tracing and is expanded upon by Chapter 4. Chapter 4 take the method and theory from Chapter 3, expands on it a bit, and applies it to the corpus of major financial legislation to see how various conditions separating the bills changes how Congress engages in idea tracing.

Getting the Message Across: Evaluating Think Tank Influence in Congress

Think tanks and position advocacy groups dominate public discourse on policy in this country, often setting the parameters of the debate, influencing the decisions of policymakers, and informing both sides of the discussion. While these organizations appear to influence a number of policies, the ways in which they exert this influence and shape politics is understudied. The overwhelming literature on interest groups deals with what Olson (1965) termed “economic pressure group organizations” or special interest groups, ones that exist to extract rents and obtain favors for a specific constituency. Public interest groups, including policy advocacy organizations, think tanks, and hybrids of the two, on the other hand, deal with the realm of information as capital. Just how they might exercise influence in the policy process is the question posed in this chapter. They must convince members of Congress that not only is their information the “best” in terms of quality, but the most relevant to any given member of Congress (MC).

Discussions of think tanks in social scientific settings have been limited to work that discusses think tanks descriptively or conceives of them as sociological phenomena detached from policy research. Rarely have they been integrated with the research studying the policy

universe, the exception for when they do assesses their role in foreign policy (Ableson 2009; McGann 2006; Stone 1996). This literature has real difficulties with the multifaceted interests and concerns think tanks bring to the table; it is far too easy to ascribe to think tanks a class conscious character and reduce their actions to their role in maintaining elite political control, seeing think tanks as the manifestation of the moneyed interests of corporations (Stone 1996). Most research has yet to consider the think tanks as institution, integrated into a larger Washington centric policy universe that is interconnected with more well understood realms (Congress in particular). Because think tanks are a hybrid of political, academic, media, and corporate organizations, it is fitting to think of them in as institutionally detached way as possible.

To better integrate think tanks into our understanding of the policy making process, I create a theoretical framework to evaluate think tank efficacy and observe the role they have on policy making, with a particular interest in analyzing how their implicit ideology and their connectedness to inside the beltway politics relate. This effect is an indirect way of observing the transmission of ideas, which is the true currency of think tanks (Medvetz 2012). The main contributions of this chapter are the development of a new measure of organizational ideology using individual ideological scores derived from campaign contributions (Bonica 2013), and the application of these scores to the study of think tank influence in Congress. The chapter presents the first steps toward the development of a theory of how ideas persists between think tanks and all other external political actors and of how these shape the behavior of members of Congress.

Defining the Think Tank

Conceptualizing what a think tank is exactly is not as simple as it may at first seem. Often times, think tanks are just defined heuristically with several high-visibility examples used. The Heritage Foundation, the Center for American Progress, and the RAND Corporation are think tanks; semi-governmental research bodies like the CBO or the Congressional Research

Service are not. The most public presentation of a think tank is the limited exposure individual policy experts get when speaking to the media about their research, so it is hard to ascertain what the public perception of think tanks is. Rich (2004), in one of the few attempts by political scientists to discuss think tanks in terms of American politics, defines them as “independent, non-interest based, nonprofit political organizations that produce, and principally rely on, expertise and ideas to obtain support and to influence the policymaking process.” This is the motivating definition for this project.

The natural place to begin a discussion of think tanks would seem to be in the broader interest group literature. Some studies of interest groups have included think tanks as a subclass of interest groups themselves (see Grossman 2012), while most generally do not address think tanks at all (see Baumgartner et al. 2012 as an example). While the interest group literature as a whole has shed a tremendous light on how these groups try to “buy access” to Congress and how they affect particular debates and roll calls (Hansen 1991; Hall & Wayman 1990; Berry 1999; Esterling 2007; Hall & Miler 2008; Evans 1996; Yackee & Yackee 2006; Mahoney 2008; Baumgartner et al. 2009), most often the agendas and policy demands they study require a constituency effect that is not found in think tanks. Work on treating interest groups as ideologically strategic movers is particularly astute and helped me think through how think tanks might behave similarly even within their own constraints (McKay 2008). While this does not limit their perspicacity in regards to think tanks, it does require a more ground-up approach to thinking about them before we can even consider more ambitious and thorough approaches that have been applied to interest groups.

Studying think tanks in isolation is then a futile exercise, because it is fundamentally impossible to understand what think tanks do and why they do it without first understanding the broader political ecology of the parties, government, media, and academy in which they operate. It is why Medvetz’s (2012) discussion of think tank “space” is very appealing, in that he defines think tanks as being conditional on the other forces driving policy; that is, think tanks are part of the “metafield” of actors trying to define and divide power in terms

of lawmaking. They are institutionally reflective of large corporations, universities, political parties, and media firms; yet they embody a space that is not adequately captured by any of them.

What is missing from these accounts is something more context driven—independence, as Medvetz rightly points out, is something of a false choice in defining think tanks in particular, the nature of the work being produced and the people being hired suggests a lot about the organization. The vast majority of the output of a think tank will be some hybrid of social scientific research and activism that address topics that are inherently political in nature and are inherently embedded in the policy process, which is consistent with what Bertelli and Wenger (2009) found about the conditions under which think tanks are founded. People with varying backgrounds in media, academia, business, and politics write these reports, and the primary logic behind them is to persuade others to the veracity of their claims. It is best then to see think tanks as something of an extension of partisan networks. They play a highly public role in legitimizing the machinations of the party they are aligned with.

Think Tanks and Policy Literature

The biggest obstacle to understanding think tanks is simply the paucity of research on them. Think tank research is absent from the political science literature, despite their prominence in media coverage of policy debates (Medvetz 2012), and although the increase in think tank prominence has coincided with increasingly polarized legislative politics, little attention has been paid to the increasingly polarized role of expertise think tanks have provided (Rich 2004). There have been no articles specifically about think tanks in the *American Political Science Review*, the *American Journal of Political Science*, or the *Journal of Politics* in the past thirty years (Rich 2004).

Studies of congressional lawmaking are probably the most fruitful home for additional research on think tanks, particularly literature that focuses on the outputs more than the process of legislating. The specific changes to the content of major legislation should be the

major orientation of the research on think tanks: research agendas devoted to a policy-centric view of Congress dovetail with the same types of substance-driven concerns that think tanks themselves have.

The policy agendas literature pays attention to the writing of legislation and is thusly concerned with the hearings, testimony, and drafting process of bill-making and not just votes (Baumgartner and Jones 2002). Where think tanks impact the policy writing process is through the information gathering and position defending parts of legislative development. Since members of Congress are devoting more and more time to campaigning (Fleisher and Bond 2004), they have less time resources at their disposal to properly adjudicate the merits of a given policy (Adler 2002). The amount of information presented at Congressional hearings is increasing overall, even as the value of any given testimony has declined, partially because of the increase in the legislative workload and partially because of the increase in aggregate sources called to testify (Adler 2002; Diermeier and Feddersen 2000).

Think tanks add to the narrative of the policy agendas literature by informing the content of testimony, which explains the seemingly paradoxical decline in legislative time devoted to lawmaking with the increase in the number of committee hearings. Think tanks are vital to the effective transmission of information in the pursuit of transforming ideas into actual legislation, since such bills serve as “vehicles” for policy ideas (Adler and Wilkerson 2012). The information transmission model of legislative idea diffusion (Wilkerson, Smith and Stramp 2015) suggests a major influx of ideas at the committee hearing level, which is where think tanks would be able to most directly exert influence. While there have been attempts to integrate a similar framework for interest groups (Grossmann 2012; Hansen 1991; Baumgartner et al. 2008), no attempts have been made to do this with non-lobbying organizations. A policy intensive theory of legislative behavior requires a role for think tanks to be comprehensive.

Expertise and Policy Equilibrium

An important element to this chapter is the development of a model of think tank interactions with Congress as a part of an explicitly partisan policymaking process. Think tanks exist because of partisan demands for information and expertise that were previously unavailable; the need for experts to defend policies that are explicitly partisan and ideological required the development of a series of institutions that requires expert level knowledge constrained by ideology. The broad presumption here is that there are elite interactions taking place at the highest levels of the policy making community that sit at the intersection of ideology, interests, and the crafting of a public policy.

There are two sides to this theory of think tank interactions with Congressional policy makers: a demand side (Congress) and a supply side (the think tanks). On the demand side, Members of Congress (MCs) view policy expertise as a means to further their own legislative ends, with think tanks providing some of this support (Bertelli and Wenger 2009). A reasonable assumption to make is that the policies they are crafting will either have largely well defined partisan predispositions that fall under already existing ideological preferences or they are new topics, ones that generally demand more information than just partisan issues would entail. Ignoring the discussion of new topics for the moment,¹ the fact that there is remarkable ideological stability from year to year amongst MCs suggests that their policy considerations remain largely stable as well—this is borne out in the policy agendas literature in that the status quo for most policy fields is partisan centric stability (Poole and Rosenthal 1997; Baumgartner and Jones 2002). The structure of congressional policymaking then is more constrained by some interaction between ideology and partisanship.²

¹ There is evidence that the rise of new topics upon which to legislate on is rather rare; with the exception of foreign policy/national security issues that sometimes arise rather suddenly, by the time there is ample pressure for Congress to move on a new topic, there has already been a saturation of partisan expertise on the market, such that there is little ambiguity about what each ideological side of a policy looks like (Baumgartner and Jones 2010).

² Not that we can readily separate the two when it comes to MCs.

This consideration is important in understanding how MCs view think tanks because it explains how they interact. The ways in which Congressmen use think tanks to further their ends is twofold: they cite them in floor speeches to bolster their own already existing policy positions or to attack the policy positions of members of the opposite party, or they call them to testify in committee and subcommittee hearings, in which case their motivation is more flexible.

My argument is that the rationale behind calling experts from think tanks to testify is directly related to a balancing logic of expertise. The expectation of most subcommittee hearings vis-à-vis the demands on policy experts are no more than a recitation of a simple ritual: I call on my “expert” to testify, you call on yours, and then the public adjudicates based on what was said. While this version of the behavior of committees is the most civically reassuring, it simply has no bearing on public cognition of policy debates (Gilens 2001; Bartels 2005), so it lacks any substantive motivation.

Think tanks are called to testify in hearings to serve the two primary functions, first as part of Congressional Oversight as a “Fire Alarm” service, and secondly they are able to provide information to Congressmen about important motivations (McCubbins and Schwartz 1984). Testimony requires substantial effort to validate the opinions expressed; it is not enough to have someone support the same policy positions that a MC does. They have to appear to be an expert as well, and they have to appear to have come to their conclusions through “legitimate” intellectual means (Adler 2002).

This model suggests an important consideration: although rarely will a specific testimony matter in the ultimate passage or writing of policy, the presence of partisan expert testimony exists as a counterbalance between the two sides. There is an equilibrium reached between both sides, in that the marginal impact of any given testimony is rather slim. If we were to assume a null case where there was no testimony, then the presence of a single expert would provide more legitimacy to one ideological side over the other. Further, if one side were to unilaterally withdraw from the expertise side of the policy fight, and therefore call no

experts to address concerns, then that would dramatically limit those MCs from that party to make persuasive cases at the committee or floor level to push policies in their direction. Recall then, concerns from Republican House members in the pre-Heritage era. They had no resources to pull from to defend their policy positions (since the American Enterprise Institute only provided research on bills after they were passed), so the compromises that were made on legislation were often less favorable than they would have imagined possible (Edwards 2001), due to a disequilibrium of partisan information and expertise valence.

The supply side of the expertise/policymaker relationship is the creation of think tanks to fill policy space. Given the ever-rising need for policy expertise, the utility that MCs gain from citing think tanks rises concomitantly as well. At the highest levels, think tanks are a collection of elite political actors who are interested in changing policy to better match their preferences. Boards of Directors of think tanks are tasked with crafting a research and advocacy agenda that is able to influence policy changes in Washington, under the constraint of staying within the ideological space of their donors (and more often than not, in the ideological space of their board members as well).

The considerations think tanks make towards policy, then, must be understood to be deliberate strategic thinking about maximizing the utility of a given think tank constrained by donor expectations and preferences. Since board members usually combine wealth with an interest in policy, their continued presence within think tanks suggest there should be some actualized returns on their investment. And if that is true, the value added by being part of a think tank is certainly worth the monetary investment they must make.

Think tanks must ultimately act to maximize their policy credibility(i.e., their propensity for producing ideological consistent work) and their overall reputation, which is their propensity for producing quality work. Both reputation and credibility play into the ultimate goals that a think tank will have influencing policy change. Pursuant to these ends, think tanks are faced with strategic tradeoffs about where they position themselves ideologically. Related to the discussion of the supply side of policy formation, think tanks saturate the market

to allow for most ideologically congruent expertise so MCs can cite them, which bolsters their position taking strength. Think tanks that are less ideological come into existence to increase the legitimacy of the policymaking process itself and to serve both sides additional policy relevant feedback.

To better understand this relationship, it is necessary to think about how think tanks relate the tradeoff of their public ideology to their reputation as a serious producer of policy research. Consistent with this tradeoff emerges the two different forms of influence discussed in this chapter: citations, which represent low cost affirmations of ideological proximity (similar to Mayhew's (1973) position taking), and calls to testify, which represent higher cost calls that require more explicit expertise than just partisan affirmation. These ideas give us two sets of very clear hypotheses. The first set has to do with the interplay between ideology and the two different forms of influence. For each think tank:

H1A: As ideological extremeness increases, citations will increase.

H1B: As ideological extremeness increases, calls to testify will decrease.

A secondary set of assumptions deals with how think tanks position themselves in the policy making network. Think tanks are only able to influence policies when they are close enough to the policy making process to get the attention of MCs before bills are written. Because of the need for proximity, think tanks value employees who have experience in the federal bureaucracy or in Congressional staffs for their ability to get the research done by the think tanks into the hands of MCs. Direct influence can be sought by tapping into networks of "revolving door" ex-Congressional staffers, former federal bureaucratic employees, and high ranking political appointees or civil servants. Think tanks are no different than any other organization that seeks to maximize social influence in policy making; these social connections are vital to gaining the access to MCs so that the organization's policy preferences can be well met. Because of this dynamic, an organization's proximity to power acts as a lowering of the informational transaction costs for MCs to engage their policy literature. Therefore, the secondary hypotheses are that for each organization:

H2A: As proximity to power increases, their citations will increase slightly and;

H2B: As proximity to power increases, their calls to testify will increase significantly.

This chapter hopes to be the first to accurately report the dynamic relationship between think tank operations, ideology, and the policymaking process.

Measuring Think Tank Ideology

I turn now to the new measurement for and procedure for the testing of these hypotheses. For this chapter, I propose an original measure of ideology for think tanks relying heavily on research done on scaling campaign contributions. The first step to understanding think tanks as policy actors would be to ground their activities in some more coherent ideological framework. While there is a some literature discussing think tank ideology (e.g., McGann 2009, Stone 1995, Rich 2004) most of these use simple designations, either made by other policy insiders, by members of Congress, or the media, for think tank ideology.

Thinking about measuring the ideology of a think tank is something that does not lend itself to an easy metric. The dominant method for measuring ideology in political science literature is the DW-NOMINATE scaling technique, which uses roll call votes to map preferences, mostly, onto a uni-dimensional scale.

What these methods all have in common, though, is that they are reserved for the study of legislators and/or survey respondents (Poole and Rosenthal 1985, 1991, 1997, 2001; Clinton, Jackman, and Rivers 2004; Poole 2000; Shor and McCarty 2011; Ansolabehere, Snyder, and Stewart 2001; Tausanovitch and Warshaw 2012; Bonica 2014). While performing a survey on think tanks might be of interest, the nature of the work they do and the limited number of people employed by each think tank would make such work more difficult from which to generalize. McKay (2008) developed a new technique to explore interest group ideology, using W-NOMINATE on interest group policy endorsements. This technique is not directly applicable for the study of think tanks since there is an explicit IRS ban on policy endorsement for think tanks, as long as think tanks are designated as 501c(3)

organizations³. McKay’s logic of deriving ideology from legislators and interest groups can be done similarly, conditional on the assumption that their preferences line up in the sense of ideological proximity. The motivating logic behind McKay’s work is tremendously helpful in thinking through how think tanks would use ideology to position themselves strategically.

My method of deriving think tank ideology is to take advantage of an exciting new dataset, Adam Bonica’s (2014) ideology scores derived from contributions made to political campaigns. Using Federal Election Committee data on contributions from individuals and organizations to politicians, Bonica used a spatial scaling method to derive single dimension ideology scores based on giving patterns and intensity. Basically, his system takes each contributor and places each of their contributions to any political campaign in a matrix that compares this behavior to other contributors and to those they are contributing too. By clustering contributions similarly to how Poole and Rosenthal cluster votes, Bonica is able to derive a uni-dimensional ideology scale that is cross-applicable to all candidates and contributors. His data includes over 40 million unique contributors, covering each election cycle from 1980 through 2012.⁴

This data provided me with the unique opportunity to derive the ideological space for a think tank. Although think tanks themselves are not allowed to donate money to campaigns, think tanks are made up of individuals who are under no such obligations. While scouring the entire campaign contributions history of people at each think tank would be valuable, there is a smaller subset of people that are most relevant to guiding the work of a think tank—and this work would be the most important manifestation of their ideology—their board of directors.

The idea that the board of directors of a think tank would have set policy goals and set ideals prior to funding the think tank, is not an assumption that is particularly daring.

³ Which all of the organizations in this chapter either are or have an associated 501c(3) wing.

⁴ The data used here were pulled from his database in the Spring of 2014: he will update his scores given the 2014 election cycle, but those updates are not reflected in this chapter.

Since they control the funding, and since many of them are already established donors to major players on either side of the political party divide, it makes sense to think of board members as investors in the policy outcomes that come from a think tank. Bonica has done work along a similar vein when comparing industries and professions, but this is the first example, as far as I am aware, of this technique being used for non profits (Bonica 2015).

The boards of think tanks are a non-accidental choice for this line of analysis as well. Being a board member of a think tank would generally suggest that a person is already political active, is most likely a serious donor, and has a vested interest in public policy concerns. Given the profile of board members in this database, it is clear that they were good candidates to have their ideology analyzed using donation patterns; they tend to give at very high rates—often times in very public ways—they tend to work in industries that are naturally very high profile to begin with, which made matching their records easier,⁵ they tend to be overt with public statements about their political preferences, etc.

Nevertheless, there are some concerns with this technique, the most practical being actually going through Bonica’s data and finding each of the board members of each think tank, given that some boards have as many as 50 members. When searching the records, it is hard to differentiate people based on the limited information included in the FEC records, especially since none of it is assiduously checked by the FEC.⁶ When doing these comparisons, it would not be uncommon to find, in some cases, several dozen people with the same name, so sifting through that and trying to determine which person listed is the one that is on the board can be somewhat time consuming. Among the think tanks in my dataset, I was able to identify around 65% of their collective boards of directors, with the only ap-

⁵ There were many corporate CEO, high profile investors, politicians, military personnel, and civil servants among those listed as members of a think tanks board. They are hardly a random selection of individuals.

⁶ There are concerns that it is a practice of major investors to obscure who they are when filing these FEC donation forms, and that they will often include misleading information about profession and state of residence (especially since many have residence in multiple states). Bonica mentioned to me that it has made matching unique contributors sometimes more difficult than otherwise would be, although his algorithm, when compared with doing the comparisons by hand, did nearly as well. But these are concerns endemic to any use of contributions data. Further discussions of this are included in the appendix.

parent systematic relationship between the members being their citizenship; since this data is derived from FEC records on campaign contributions, it limited the scope to only United States citizens.⁷ This number is similar in scope to what Bonica finds in his own papers (Bonica, Johnston, and Chen 2015; Bonica 2014).

I was thus able to create an estimate for 133 think tanks, all of which appeared in Groseclose and Milyo’s (2005) study of think tanks in Congressional and media citations. They selected their think tanks from the *Think Tanks Directory in America*, and from that selected the ones that were cited in the *Congressional Record* (Groseclose and Milyo 2005).⁸

This chapter uses these newly generated scores and compares them with the only other reliable scoring method for think tanks currently in circulation—Groseclose and Milyo’s (2005) method that uses citations in the Congressional record to map think tank scores to Congressmen to whom they give ideology scores using Groseclose, Leavitt and Snyder’s (1999) method. Their scale is derived from a weighted, yearly-recomputed measure that uses scores from the liberal PAC, Americans for Democratic Action (“ADA”), and assigns legislators a score along a zero-to-one hundred scale, with zero being most conservative and 100 being most liberal. Their paper, which used their ideology scores of think tanks to assess bias of media sources, is the only previous attempt to measure think tank ideology. But it runs into a very important problem; their scores cannot be used to predict anything relating to Congress, because they derive their estimations of think tank ideology from the Congressmen citing them, making them hopelessly endogenous for this purpose.

I take the Bonica scores for the members of the boards of directors and I generate a simple arithmetic mean for the organizations board. This score is referenced to as their

⁷ The only think tanks I was unable to get enough members to include in my dataset were the Third Millennium Foundation and the International Policy Institute for Counter-Terrorism, both of which were largely made up of foreign nationals. For the boards generally, I only included members in my dataset if I could verify that their known giving patterns (from the FEC website) matched their scores in Bonica’s data.

⁸ Limiting the data to think tanks featured in this guide could be potentially bias some aspects of the results, insofar as it will be left censoring on the distribution of think tank influence. However, limiting the universe to think tanks prominent enough to make it into this guide gives us more comparable units of analysis.

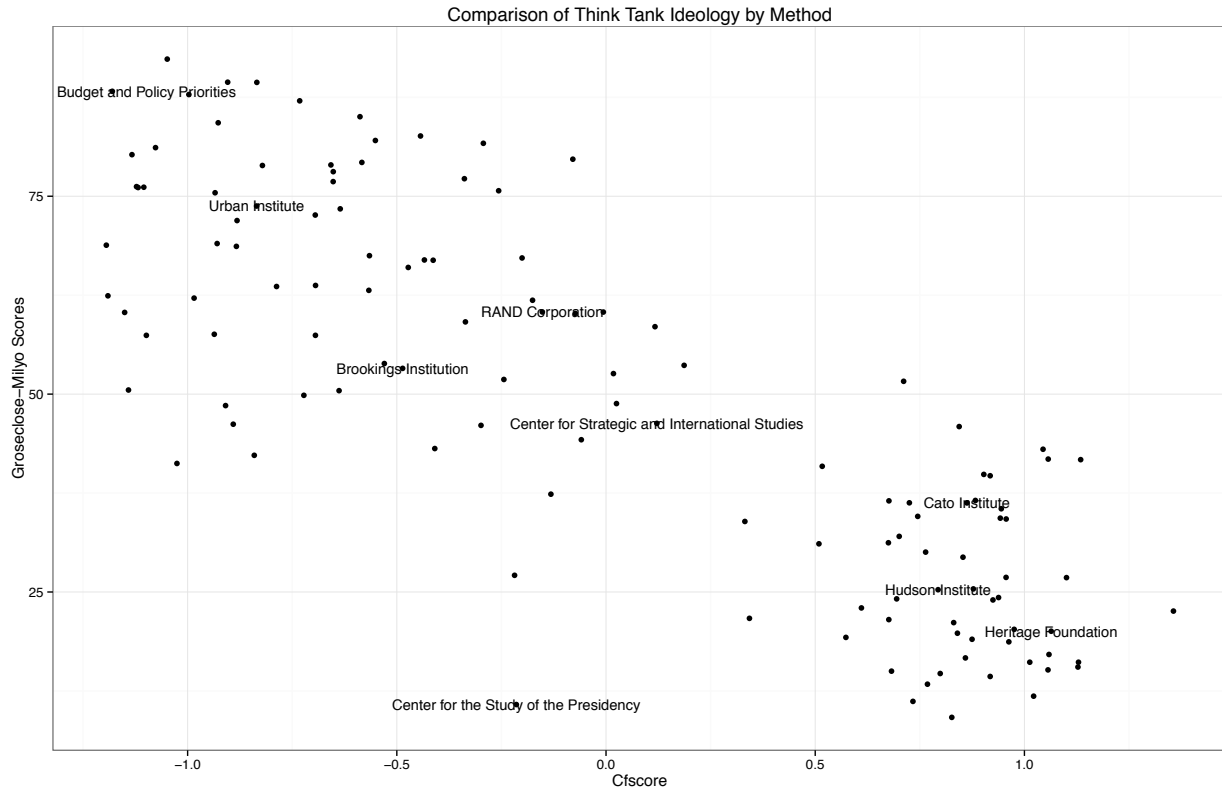


FIGURE 2.1: Comparison between Groseclose-Milyo Scores and CFscores for Think Tanks

“general ideology scores.” Since most organizations have large enough boards where this type of technique is reasonable, it is perfectly reasonable to take a collection of individual ideology scores and aggrandize them at an institutional level.

The first thing that needs to be done is to compare these scores against the Groseclose-Milyo scores (which run from 0 to 100, with 100 being most liberal and 0 being most conservative), as well as analyze them using what we already know about think tanks. Looking at Figure 2.1, we see a very strong relation between the two sets of scores; there is a 0.82 correlation between the two variables.⁹

⁹ The most significant difference between the two being two distinct think tanks, the Center for the Study of the Presidency, which GM score as highly conservative and I score as neutral, and the Christian Coalition, which GM score as moderately conservative and I score as the most conservative. Since the Center for the Study of the Presidency is a non ideological think tank by design, I am highly confident that its true ideology makes more sense in my scale than in GM. Similarly for the Christian Coalition, since it is the think tank founded by Pat Robertson, it is most likely toward the extreme of the conservative movement, not near the center. Both of these, and the few other reasonable deviations between the scores, are mostly caused by the

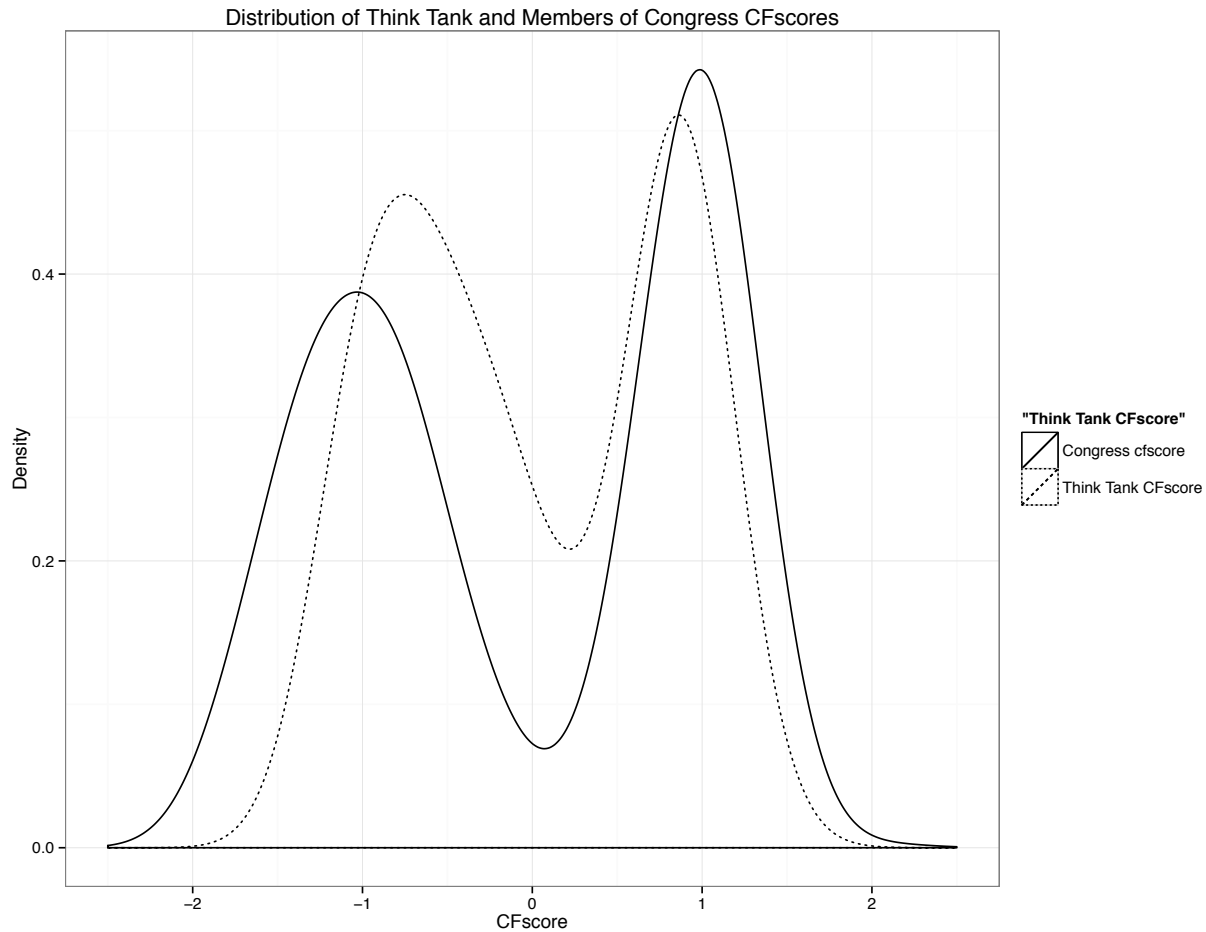


FIGURE 2.2: Distribution of Congressmen and Think Tanks by CFscore

This is borne out in the rest of the data as well. The correlation between the two sets of scores for think tanks receiving under 20 citations in Congress is only 0.71; for those above 20 citations, it is 0.89. Bonica's data is normalized to have the center be right around 0, with a scale of -3.5 to 3.5, but most of the data for candidates exists between -1.5 and 1.5. For a point of comparison, the two parties' medians in Congress currently are located around slightly more than -1 and slightly less than 1 for Democrats and Republicans respectively. The CFscores, though, are most useful for their ability to test theories about influence and citations in Congress.

paucity of citations they received; the Center for the Study of the Presidency and the Christian Coalition were only cited one and three times respectively.

In Figure 2.2 above, we can see that the distribution of Congressional ideology and think tank ideology is very similar, suggesting a high degree of face-validity in the idea to measure these two in a similar fashion, and that there are definitely some forces shaping the ideological ecology of think tanks. Further discussions of these original scores are found in supplemental appendices.

Estimation Strategy

To best understand how think tank ideology directly impacts Congressional activity, we need to look at two distinct dependent variables for this study: Congressional citations and Congressional testimony. Congressional citations refers to the frequency with which a specific think tank is mentioned in the *Congressional Record*, and Congressional testimony refers to the number of times a person who is identified as being affiliated with a given think tank is called to testify in either a committee or subcommittee hearing. This empirical strategy can be thought of as a form of the nonequivalent dependent variables design.

These are obviously not perfect measures of influence in Congress, but they are excellent starting points for thinking about who is, at a minimum, making themselves prominent amongst the policy writing community. Think tanks desire this sort of preeminence, so using these two metrics as related but distinct measures of influence is a rather straightforward way of getting at an indirect measure of influence.

To predict think tank citations, I used a Bayesian Poisson Regression Model with weakly informative priors, which uses a Bayesian Estimation Maximization (EM) algorithm that takes estimates of the coefficients from the frequentist version of the regression, then iteratively comparing estimates of the coefficients operating off of a simple set prior (in this case a Student's t-distribution), where each of the coefficients is estimated with a normal distribution, and the scale is assumed to have an error term distributed by the t-distribution. Given that the DVs are both counts, a Poisson or Negative Binomial regression of some form is preferable to traditional OLS. For identifiability, I use a weakly informative prior for the

means and error terms of the coefficients that sets the value to zero. The nature of this type of Bayesian regression model allows for simple interpretation in a frequentist framework, while maintaining the inherent flexibility and power of Bayesian methods. The best reason to use a Bayesian model here, even one with just a weakly informative prior, is the nature of my data; Bayesian approaches have been shown to be more consistent amongst lower sample size, and higher complexity models (Gelman et al. 2008). There is also evidence to suggest that these types of Bayesian models are less susceptible to over-dispersion as a means of biasing the error terms (Gelman et al. 2008). The parameters of the models are as follows

$$Y_i = \beta_0 + \beta_1 \text{Employees}_i + \beta_2 |\text{CFscore}_i| + \beta_3 \text{Age}_i + \beta_4 \text{Lobbyists}_i + \beta_5 \text{Expend}_i + \epsilon_i; y_i \sim \text{Poisson}(\lambda_i)$$

where

$$\lambda_j \sim \text{exp}(x_j, \beta_j) \text{ and } \beta \sim N_k(b_0, B_0^{-1})$$

which describes the Poisson regression.¹⁰ In each case, Y_i is either the count of Congressional citations for a think tank, or their testimony counts. Each of the other variables are measured as follows.

The first independent variable listed is the “Employees” variable. This is a metric designed to capture elements of “embeddedness” of a given think tank to the policy making process. Employees is the measure of the number of employees at a think tank that have previously worked in either the federal government, the congressional bureaucracy, or are (or were) registered lobbyists. While this will have some overlap with the Lobbyists variable (more on that in a bit), this variable is distinct and is designed to measure how a think tank pulls its employees from other parts of the policy-writing process. This is either an incidental measure of embeddedness, in that the think tanks just happen to employ people with this very specific set of backgrounds, or that they deliberately select employees with this type of background with the explicit purpose of increasing their influence by using the

¹⁰ The classic parameters of a Bayesian approach to the student’s t-distribution

employees connections. This variable ranged from zero employees to fifty employees, with the majority having four or fewer employees.

My second independent variable is the absolute value of the CFscore of the think tank. I use the absolute value because I'm not necessarily interested in the direction of the ideology of the think tank but the "extremeness" of it (i.e., the distance from zero for the ideology scores will correspond to some measure of how far from neutral they have placed themselves ideologically). By taking the absolute value, it should eliminate structural differences caused by the ideological space occupied by the two parties and should more or less eliminate any undue bias from changes in party control of Congress during the time frame. This variable ranged from zero to 1.35, which essentially covers the full scope of CFscores (in absolute values).

The next variable is the age (in years) of the organization in 2005, the final year the data cover. The ages range from five years, to 134 years, with a median of 27 years. Following that, we have the Lobbyists variable, which is a dummy variable for whether or not the organization either hires lobbyists themselves or has a related organization that engages in lobbying. For the most part, this was very easily determined from their website, as several of the organizations had more famous wings that were lobbying organizations, designated as 501(c)4 organizations, while the foundation or think tank attached is a 501(c)3 organization. The final variable is a control for overall size: the total expenditures for the think tank in the earliest fiscal year available or in 2005, whichever was attainable. This variable had a tremendous amount of variation, with the smallest organization only having a budget of \$54,571, whereas the largest had a budget of \$275,076,526, with the mean organization having a budget of \$19,240,000 and the median having a budget of \$4,819,000. This should be an ample yet strong control for the variations in size and monetary power of an organization, and, combined with age, should be good bulwarks of non-influence based measures of prominence, and should help clear the analytic picture.

Results

Table 2.1 shows the results of the two different regression models, with different variables excluded to better illustrate the effects of our independent variables of interest. The first group of models deals with Congressional citations and the second group deals with Congressional Testimony. The first model for each category began with just the “CFscore” variable so that we could make baseline estimation without just the controls. The second category adds in Employees, and the third regression adds in other controls. Staggering the variable selection should make it clear that the covariates are not selected to just enhance the main finding, but for their theoretical contribution.

The first thing to note from these models is the difference in the N’s. From the 133 think tanks in the database, I was unable to get expenditure information for 4 of them, and I was unable to, because of missing data, to get the boards for another 4 think tanks, explaining the (slight) discrepancy between the observations from the models.

The biggest conclusions from the regressions in Table 2.1 are that the two independent variables of interest, the number of registered former federal employees a think tank hires and the absolute value of their ideology score, are both significant in every case. The effects are consistent: for Congressional citations and for Congressional testimony, the more former federal employees in your organization (i.e., the greater your proximity to power), the more prominent your think tank. The covariates generally were significant in the direction one would expect, with the noted caveat that *having lobbyists did not increase the propensity for an organization to have someone called to testify*. Thus, these findings support H2A and H2B in the most direct manner.

The most interesting finding here, though, was the effect that the absolute value of CFscore had on citations and on testimony, as in H1A and H1B. For citations, the relation is fairly strongly positive: the more extreme your organization is ideologically, the more it gets cited. For testimony, however, the relationship is directly reversed: the farther

Table 2.1: Estimating Citations and Testimony in Congress

	<i>Dependent variable:</i> <i>Citations</i>			<i>Dependent variable:</i> <i>Testimony</i>		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	4.06*** (0.03)	3.67*** (0.03)	2.56*** (0.04)	5.50*** (0.02)	5.05*** (0.02)	4.05*** (0.03)
CFScore	0.37*** (0.03)	0.39*** (0.03)	1.22*** (0.04)	-1.11*** (0.02)	-1.33*** (0.03)	-0.41*** (0.03)
FedEmployees		0.05*** (0.00)	0.03*** (0.00)		0.06*** (0.00)	0.05*** (0.00)
Age (in 2005)			0.01*** (0.00)			0.01*** (0.00)
Lobbyists			0.04*** (0.01)			-0.00 (0.01)
Expenditures			0.01*** (0.00)			0.01*** (0.00)
<i>N</i>	125	125	125	125	125	125
AIC	14975.98	12528.72	10006.27	27202.28	17703.00	10543.71
BIC	14998.61	12562.66	10074.15	27224.90	17736.94	10611.59
log <i>L</i>	-7479.99	-6252.36	-4979.13	-13593.14	-8839.50	-5247.85

Standard errors in parentheses

† significant at $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

your organization's ideology is from neutral, the less likely (or less frequently) you are to be called to testify. The extant different directions for the key independent variable for the two dependent variables are what lend itself to the nonequivalent dependent variable design, and allows us to think of it as a test of the two provisions. This appears just as the aforementioned hypotheses claim. But parsing out the magnitudes of the direct effects here, even if the baseline effect confirms H1A and H1B, is impossible with just a simple regression table.

To better adjudicate the direct effects of these two variables, I created two figures that measure the estimated citation and testimony counts for organizations along the range of ideology scores and along the range of number of connected employees. Figures 2.3 and 2.4 show this, with the lighter shade representing citations and darker shade representing testimony.

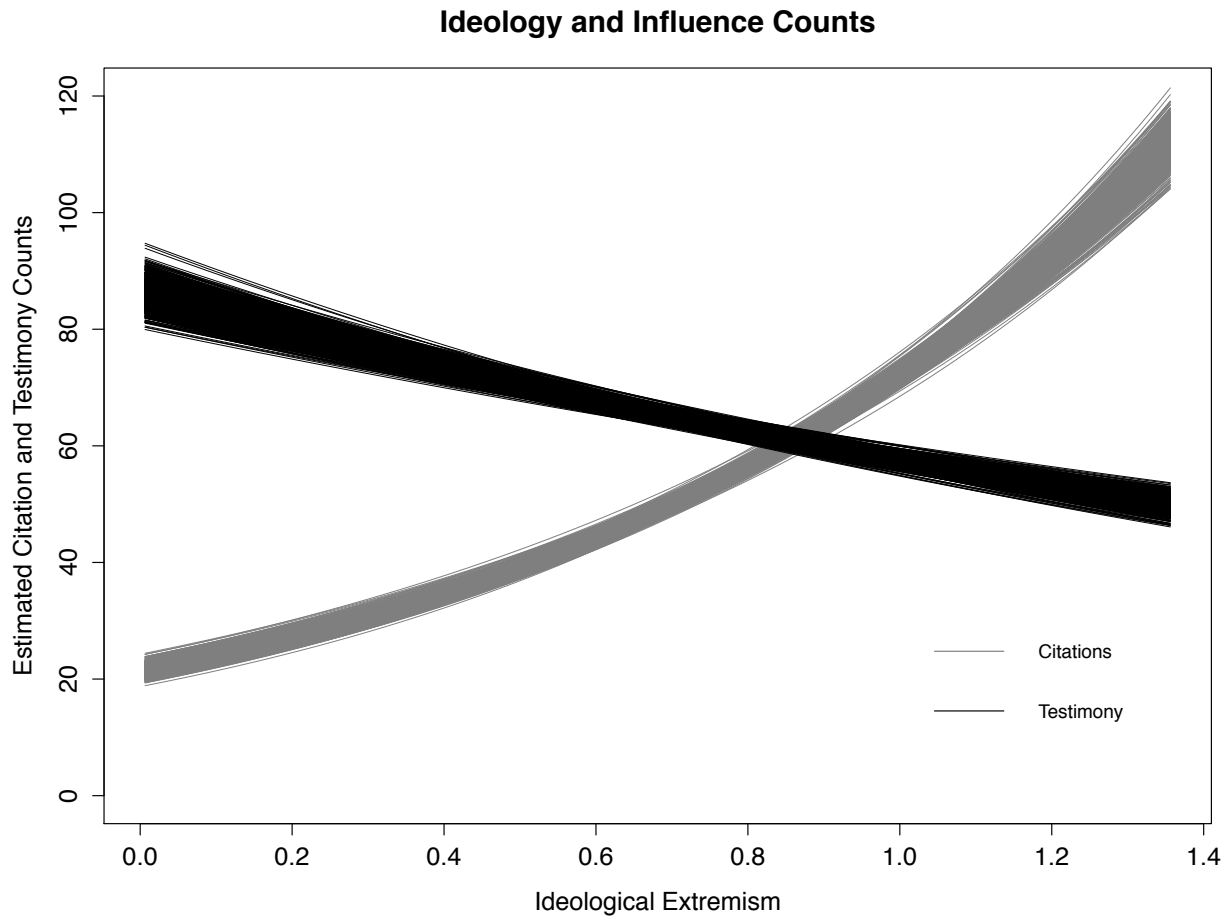


FIGURE 2.3: Estimated Citations and Testimony holding expenditures, age, and connected employees at the median, with no lobbyists.

Both of these figures take the posterior distribution of the beta coefficients, multiplies them by basic inputs for the independent variables (in this case, selecting the median values for age and expenditures, and assuming no lobbyists) excluding the variable of interest, which is then allowed to vary along its range of values, and transforms them through their link function (the log function). So the ideology graph runs from zero to 1.35, and the connectiveness graph runs from zero to fifty.

This effects graph allows for easy comparison of the effects of changing the variables of interest on the dependent variable in real terms. In the ideology graph, for every additional 0.2 value of the absolute value of your CFscore you receive around an extra 20 Congressional

Connectiveness and Influence Counts

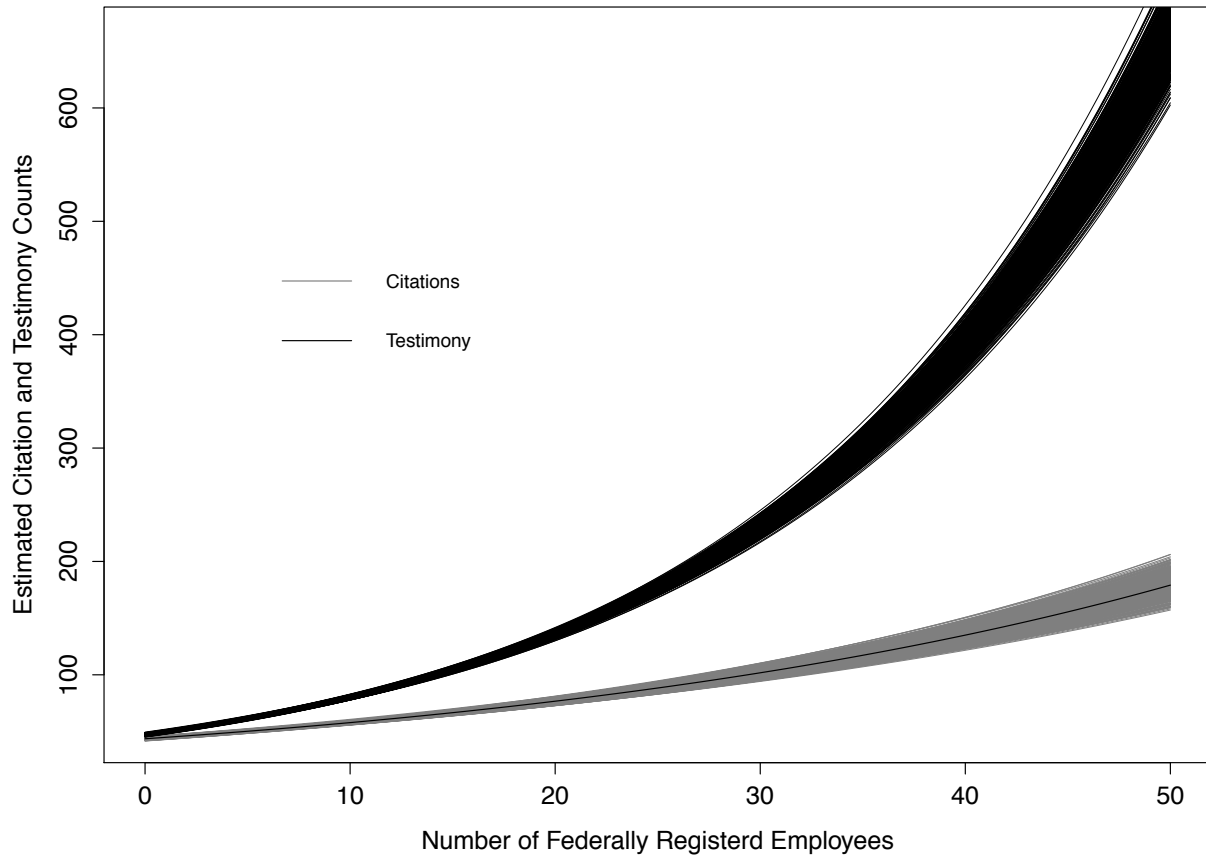


FIGURE 2.4: Estimated Citations and Testimony holding expenditures, age and connected employees at the median, with no lobbyists.

citations; the same unit change in ideology leads to a loss of around 10 calls to testify. Theoretically, if one organization were to value citations and testimony at an equal rate, they would extract the most utility by placing their ideology right around 0.8 magnitude from zero, or just slightly more neutral than both party extremes. This graph is an excellent illustration of the strategic trade-off that think tanks have to pursue to maximize their influence, as they cannot force their ideology too far to the extreme so that they are no longer considered “experts” in their field, yet having some ideology makes a think tank more serviceable as a citation for MCs. These findings completely affirm hypotheses H1A and H1B of this chapter.

The most interesting point of all here though is that *the two estimated counts graphs intersect right around 0.85-0.90, which is right around the median for each of the parties in Bonica's database.* The fact that the convergence occurs right around the median of the two parties suggests that there is definitely some credence to the idea that think tanks strategically position themselves ideologically to maximize their effectiveness on both dimensions, leading to a partition-style equilibrium in think tanks.

The other finding is with the second variable of interest, the number of federally connected employees. For Congressional citations, the effect is moderately powerful: for every 10 more federally connected employees, a think tank should expect 20 to 30 more citations. The effect also strengthens when you get beyond the 10-20 range of federally connected employees. Where this effect is even more pronounced, though, is with the testimony data. Depending on how many employees you have to begin with, every additional 10 federally connected employees increases your call to testimony by anywhere from 70 to 200 additional calls to testify, and the effect increases in magnitude as you increase your total number of already existing employees. The magnitude of this effect is pronounced, and suggests that the overall proximity to power an organization has is inseparable from ideology. Remember, this effect persists when we hold the size of the organization constant, so this is not just a story of money or total employees. This directly affirms the predictions made by hypotheses H2A and H2B as well. While this effect only has a tepid impact on the role of citations, the forcefulness of the result on Congressional testimony suggests an even more important insight into the black box of policymaking and bill writing.

Conclusions and Future Work

These findings suggest several very important things about the relationship between think tanks, ideology, and the development of public policy. First of all, there is definitely a strategic consideration to be made when framing the ideology of an organization that deals with the policy world. Consider the distributions of CFscores for think tanks and candidates

for Congress again (Figure 2.2). The almost perfectly congruent distributions of the ideology scores suggest a deliberate matching done by the think tanks to match MCs, such that they will position themselves in such a way to benefit the MCs most directly. This relationship exists to reinforce the MCs already-existing policy preferences with at least name-recognition level support by the think tanks, in exchange for more direct impact on policy and more prominence. There is clear evidence that think tanks strategically position their ideology such that they are most useful to MCs who have a demand for legitimate expertise to bolster their positions.

But this relationship is not simply one of think tanks matching ideologies without trade-offs. These organizations have to temper their ideological shifts with the perception of legitimacy. Expertise, then, is extended by their placement as an ideologically neutral organization. This highlights the complex role that ideology and expertise play in the pursuit of influence in Congress. Further, there is no reason why this relationship wouldn't extend to other organizations and people that are used in the policy making process to improve the ease of legislation, like lobbyists or academics (to look at two groups on opposite ends of the advocacy-expertise continuum).

Secondly, within the context of a policy-making framework, having an organization be more connected greatly improves the reach of its influence. Not only does this improve influence, but also it at least seems like it entirely predicts the ability to get called to testify on policy issues and supplants the basic reputational effects on getting public attention. This finding is in particular an exciting finding because, in no small part, it is largely unexplored. The number of employees who are on the registry of former federal employees is far from the best measure of proximate "centrality" of a given organization.

A more parsimonious approach to this question could be bridged with the burgeoning field of social network analysis, particularly the ideas of the Extended Party Network. A potentially very lucrative avenue of future research would be to try and model the social interactions of members of a think tank with people associated with Congress (either MCs,

their staffers, or members of the bureaucracy) or with the federal government. A potential project would look at the relationship between Congressional attendance at events held by the think tanks and future calls to testify in front of Congress. One could even model them at the individual level and try and track scholars to Congressional staffers to MCs. This, at a minimum, would get us to a more direct measure of how much informal interactions are had between think tanks and MCs.

Another way to further this line of research is to deconstruct the notion of what influence is. A line of research pursued in later chapters, would be to define “influence” not only in terms of who is being called to testify before Congress or who is referred to in Congressional speech (two measures that probably better approximate elements of buttressing already existing position taking than do original influence), but in the language of both the Congressional debate on issues and on the bills themselves. A Text-as-Data approach would potentially allow for tracing of policy ideas directly through the written work of think tanks into the bills themselves, and may allow for easier tracking of where exactly policy ideas originate (Wilkerson, Smith and Stramp 2015; Blei et al. 2003).

A final extension to this chapter would be to take the method of estimating organizational ideology, the averaging of Board of Director CFscores, and apply this to other equally political organizations. Potentially interesting topics include determining the role that organizational ideology plays in obtaining federal grants for specific programs, the funding of controversial projects (ideology scores of foundations themselves would be interesting), and potentially coming up with a way to give ideological scores to interest groups and try to estimate probability of the passage of a certain bill or bills based on the ideology of those external groups supporting or opposing them.

These extensions would further demonstrate the relevance of these findings, and reinforce the theoretical case for the study of external influence on the policy writing part of Congress. Further, these extensions would do well to integrate studies of policies with studies of changing language and of social interaction, all while informing a discussion of the changing role

of expertise in the policy sphere. Connecting the studies of language, interest, ideology, and policymaking is certainly of the utmost importance to the continued study of politics.

Using Ideas as My Maps: Exploring the Flow of Ideas From Hearings to Legislation

Introduction

Where do ideas come from? Ideas are clearly important to understanding the behavior of any legislature and are essential to actual legislating. In drafting legislation, Congress hears diverse and often conflicting ideas for what should be written into law. What determines which ideas affect how laws are drafted?

In this chapter, I will take these computational approaches to the study of language and develop a new method to trace the flow of ideas in legislation, and I introduce a novel way of approaching both legislative politics specifically and politics and policy generally. This approach is designed to be simultaneously straightforward in its implementation and precise in its ability to identify specific ideas as they develop. In many ways, this represents a more formal and mechanized form of idea tracing that would go on in more qualitative work and should be thought of as an extension of these approaches, not a replacement.

This project addresses one of the fundamental questions to the functioning of democratic governments: does the legislature really represent the best interests of the public? While

this project does not deal with voter preferences, it does address the content of legislation being passed and, in doing so, speaks to the basic ways in which the legislature implements public policy concerns. The creation and diffusion of policy ideas in the legislative process is fundamental to understanding exactly the ways in which the legislature is writing and implementing the laws that are part-and-parcel to the successful implementation of democratic will. If we care at all about how legislators accurately create policies that both represent the public at large and do their best to improve the general welfare of the nation, then we have to care about the substance of the bills being passed and the idea-generating process behind them.

As previously discussed, there exists very little literature on the role that ideas play in the legislative process. Broadly speaking, most of the attention paid towards legislation occurs in a more indirect way, focusing more on the bill as an element of the legislative process, as a tool of agenda control by party leaders, or as something voted on in aggregate to scale. Generally, legislative scholars study ideology, voting patterns, agenda setting, and a multitude of other themes, not necessarily focusing on what the legislation is made of (Shepsle and Weingast 1987; Krehbiel 1998; Poole and Rosenthal 2000; DeGregorio 1999; Cox and McCubbins 2005; Baumgartner et al. 2009). Interestingly enough, even research that is devoted to studying legislative efficiency and shifts in policy preferences of legislators do not devote much attention to the rich policy substance of legislation (Mayhew 1991; Krehbiel 1998; Binder 2003; Clinton and Lapinski 2006; Maltzman and Shipan 2008; Berry, Burden and Howell 2010).

While one would expect the literature on legislative effectiveness to tie in nicely with the policy relevance of legislation, most of the research along these lines gauge effectiveness in terms of bill progress, which only incidentally focuses on the substance of the bills themselves. (Volden, Wiseman and Wittmer 2013; Hasecke and Mycoff 2007; Krutz 2005, Anderson Box-Steffensmeier and Sinclair-Chapman 2003; Ainsworth and Hanson 1996; Schiller 1995). Similarly, while my perspective on idea transmission is very similar to the vast policy diffusion

literature, this study focuses more on ideas specific to a particular piece of legislation and does not overlap with much of that literature (Berry and Berry 1990; Volden, Ting, and Carpenter 2006; Shipan and Volden 2008, etc.)

Where the content of bills would make the most sense is in the policy agendas and agenda setting research, which portrays changes in policy as a process in which policy “entrepreneurs” are able to capitalize on legislative “windows of opportunity” driven by the major agenda shifting events (Walker 1977; Kingdon 1995; Baumgartner and Jones 1993). The problem here is that the sources of the ideas in legislation are largely event driven, and the literature assume that ideas must exist in the legislative ether from which Congress samples from in order to reinforce their priors on an issue.

These studies all focus on an element of legislative decision-making but do not devote much attention to the process by which ideas are introduced into the legislative process. The policy agendas literature is the closest body of work concerned with the hearings, testimony, and drafting process of bill-making and not just votes (Baumgartner and Jones 2002), but the information gathering and position defending parts of legislative development are not given separate attention. Since members of Congress are devoting more and more time to campaigning (Fleisher and Bond 2004), they have less time resources at their disposal to properly adjudicate the merits of a given policy in a vacuum (Adler 2002; Curry 2015). There is evidence to suggest that the value of information presented at Congressional hearings is increasing in the aggregate, while the value of any given testimony has declined, partially because of the increase in the legislative workload and partially because of the increase in aggregate sources called to testify (Adler 2002; Diermeier and Feddersen 2000).

Chapter 3 is important foundational work establishing the validity of a method and approach that will be utilized again in future work. I implement a variation of the topic modeling algorithm for generalizing topical trends about texts for the testimony surrounding the passage of the Dodd-Frank Wall Street Reform and Consumer Protection Act. Once these texts are classified by a simple clustering method on the posterior probability matrix

generated by the topic model, I use a text reuse algorithm (adapted from a popular form of plagiarism detection software) to track instances where identical text appears in a section of a bill that first appeared in testimony (Bloomfield 2012). This chapter not only addresses specific sources of influence on the writing of Dodd-Frank, but it also introduces a new methodological approach to the exploration of idea transmission.

Ideas in Politics

Ideas are an obviously important topic in politics that is rarely addressed directly by political scientists. The basis for studying ideas is largely not agreed upon—indeed, the key concepts behind what ideas are remain hard to parse out. It is most commonly discussed in international relations literature as part of larger theoretical arguments, focused on the constructivist school of thought (appropriately referred to as the “ideational” school of thought). This is posited as a direct contrast to traditional approaches to international relations that deal with power and institutions as *prima facie* facts and argue that the ideas behind them are equally important (Checkel 1998). This conceptualization, while a meaningful counterbalance to traditional approaches to political science, is best thought of as a compliment to traditional approaches in non-international relations domains.

My conceptualization takes ideas as an important offshoot of information, and instead of acting outside of normal political conditions, is instead part and parcel of rational behavior. An idea, then is a subset of information—it requires information to take hold but is itself more than just information. In this approach, I assume that ideas are more abstract and personal types of information. This means that it is some concept or fact that is acted upon by someone perceiving it. This is similar to the notion of “conceptual blending,” which argues that the only way we understand abstract concepts is by combining concrete concepts that we understand better. I argue that ideas are a realization of the conceptual blending (Turner 2001; Mackay 2003; Fauconnier and Turner 2008; Turner 2014). Conceptual blending builds off of the notion that all forms of higher level human cognition derive in at least some

small part from representational language and are based on the implementation of practical metaphors. Consider how we think about what an “ideal point” is, to choose perhaps the most discussed construct in American politics. We use the spatial metaphor to illustrate that a persons ideology exists relative to another person; we ascribe notions like directionality and dimension to this as if these are meaningful constructs outside of the spatial metaphor. Even the very idea of a “left-wing” and a “right-wing” pull from multiple metaphors. We blend together the ideas of spatial dimensions to politics with historical language about physical wings to represent location. By combining two visuals we understand already—left-right spatial dimension and physical wings in a hall—we create a meaningful intellectual representational language that delineates a series of ideological positions; we create the idea of a left and right wing ideology.

For this chapter, I argue that ideas absolutely can and indeed must be transmitted like information. For this reason, we can use methods that look at specific segments of text as the basis for ideas. Indeed, the reuse of specific words or phrases can indicate that the speaker is sharing in all of the necessary meta-context and thought behind the ideas in question. Thinking of how ideas are transmitted as being parts of an information transfer process allows us to model idea transmission using frameworks that are robust to complex interactions, including formal representations like hearings. Information theory gives us the idea that this transmission is itself reproducible in a formal setting.

Theory

With these situations clearly laid out, there are some implications for the linkages between ideas and legislation that need to be made before a general theory about ideas can be properly claimed.

The process that legislation goes through before testimony can begin is rather straightforward. A bill is proposed, presumably with ideas present in the text. This bill is then assigned to a specific committee or committees, where it is either rejected or assigned to

relevant subcommittees. The heads of the committees and subcommittees then call in experts to testify in their respective hearings in order to perform any one (or more) of three tasks. First, they address specific concerns as representative of an industry or constituency who desire Congressional action in order to ameliorate any concerns. Second, they speak to expertise on a given issue that would help in the writing of important legislation. Or third, they are called to testify as part of Congressional oversight, and the testimony is partially a public excoriation or pseudo-trial (McCubbins and Schwartz 1984; Hall and Wayman 1991; Epstein and O'Halloran 1995; Balla and Wright 2001). These motivations determine the landscape of those who are called to testify, and the time demands of hearings in Congress, which are the determinates of the legislative agenda.

Using the concept of a “policy idea” adapted from Adler and Wilkerson (2012), I argue that it is through the selective choice of expert testimony made by committee and subcommittee chairs that determine what policy ideas make it into the domain of legislative discourse; they either use it to reinforce ideas already present in early drafts of bills or to introduce a potential new set of ideas to make the legislation better fit their agenda. Since bills can be thought of as “vehicles” for policy ideas (Adler and Wilkerson 2012), and according to that “many if not most important bills are largely drafted by committee staff under the supervision of the committee or subcommittee chair” (Evans 1991; Kaiser 2013; Wilkerson, Smith and Stramp 2015). It is not unreasonable to extend this to say that MCs define the parameters of what policy ideas to sample from by their choices of who to call to testify. Broadly speaking, legislation is a product of ideas transmitted through the strategic engagement of policy experts through committee hearings/testimony.

The nature of transmitting ideas from people testifying requires a model for understanding how Congressmen are able to adjudicate which information from people testifying is reliable and will lead to achieving the public policy goals they desire. This involves a hierarchical process, in which those who are called testify are not in fact selected randomly but are systematically selected by the leadership of the subcommittees covering the issue space

in question.

While there will certainly be a variety of different sources of variation in determining who gets called to testify, there are, presumably, consistent things that predict who, once called to testify, influence the language of bills. My initial thoughts on how to think about this is to adopt elements of the Lupia-McCubbins persuasion model, in which voters actively engage in the process of learning about politics through their evaluation of a third party “speaker” acting as a go-between for voters (the principals) and the government (the agents) (Lupia and McCubbins 1998). Their relation of an information exchange for voters uses the principal-agent framework to better understand how voters, with limited interest and time resources, are able to adjudicate properly what to do with regards to voting. They rely on external speakers, who they evaluate at an individual level, to rely pertinent information that they require to make informed decisions. The process through which voters evaluate which speakers to rely on seems to be very similar to the logic MCs employ to evaluate the quality of testimony before committees and subcommittees.

The model for how MCs incorporate ideas presented at committee hearings would rely on a framework like this: The principal (MC) must perceive the speaker to be trustworthy, and the principal must perceive the speaker to have the knowledge they desire. Absent external forces, persuasion requires perceived common interests, which is most directly captured by who calls them to testify (and other signals of partisan affiliation) and perceived speaker knowledge/personal character. In the presence of external forces, these requirements can be reduced. External forces can be substitutes for common interests.

The standard external forces in this case are: costly signals (any personal stake the speaker leverages here, potentially for testifiers something akin to a reputation effect); a penalty for lying (i.e., it is a federal offense to knowingly lie to Congress, although its threat is probably more minimal); and the threat of verification, which is probably best manifested through competing expertise from other speakers (Lupia and McCubbins 1998). What is understated here, and is essential to this theory, is that the process through which

a speaker's credibility is even considered by Congress is contingent on the nature of what they are speaking of. Expertise in this regard is understood to impart information for highly complex issues in such a way that Congress cannot ignore the outcome; the combination of complexity and issue urgency will be highly predictive of Congressional borrowing of expert language in bills. Policy complexity is conceptualized similarly to how Elgie and McMennamin (2005) discussed it, insofar as it refers to policies that require expertise to understand and implement, or, as they summarize it, "the greater the issue complexity, the greater the need for policy experts" (Elgie and McMennamin 2005 pg. 545). While the theory outlined in this chapter does not rely on a reductionist definition of policy complexity, it does find evidence of exactly that relationship.

This framework should enable predictions about whose ideas actually make it into the final language of bills and under what conditions we can expect different speakers to influence the language of a particular piece of legislation. My theory broadly is that there is a multilevel selection effect at hand, that the people who are brought to testify before Congress are those who have publicized their own research with the most force, including, but not limited to, promoting their ideas through having events inside the beltway on the topic of interest. Thinking again about the "Police Patrols" versus "Fire Alarms" model of Congressional oversight, these researchers introduce a third type of testimony, designed to aid in the writing of legislation. When determining who is called and who influences the legislative process, different expectations need to be made, given the role that Congress would be playing through testimony. It is important to note that this framework of expert/speaker evaluation only holds true when Congress is expecting the outcome of the testimony to end with concrete legislation. It would not hold true, or at least hold together as constantly, when their primary role is oversight. It is also important to note that any findings here necessarily could not find the origin of ideas, insofar as that is even possible, and instead limits itself to discussing ideas that are already permeating the ideated ether, namely ideas that are developed enough to be acted upon in Congress. This theory offers no major expectations about the conditions

under which ideas are developed: only the conditions under which Congress listens to already existing ideas.

Methods

Chapter 3 utilizes two distinct text-as-data approaches to sort and analyze textual overlap between the particular testimonies during the Dodd-Frank hearings and the language of the final bill itself. To sort the testimony and bill sections into meaningful categories (to insure that the text-reuse is not just picking up on non-substantive language), I use a topic modeling algorithm, fit to testimony and bills jointly, then apply a text-reuse approach to find exact textual overlap between testimony and bill sections.

Structural Topic Model

Probabilistic topic models are a collection of algorithms that aim to discover the hidden thematic structure in large archives of documents. Topic modeling is a fundamentally Bayesian approach to discovering the underlying logic and structure of documents by analyzing *only* the words of the original texts, which in turn is used to determine the themes that run through them, how those themes are connected to each other, and how they change over time. Topic modeling relies on detecting the hidden structure in texts, in order to tease out the ideas behind them, while assigning posterior probabilities that any given document contains any generated topic (Blei, Ng, and Jordan 2003; Blei and Lafferty 2007; Blei 2012).

Topic modeling is a way of automatically discovering topics that these documents contain. For example, suppose we want to find two topics, A and B. A topic-modeling algorithm assumes that the author of the documents set out to write these documents with n words (with the words being distributed by a Poisson Distribution) that contains a set topic mixture of the two different topics (distributed as a Dirichlet Distribution), and then using a multinomial distribution to pick words within each topic. The model then reassesses from the documents to find the underlying set of topics the author wanted to convey and does so

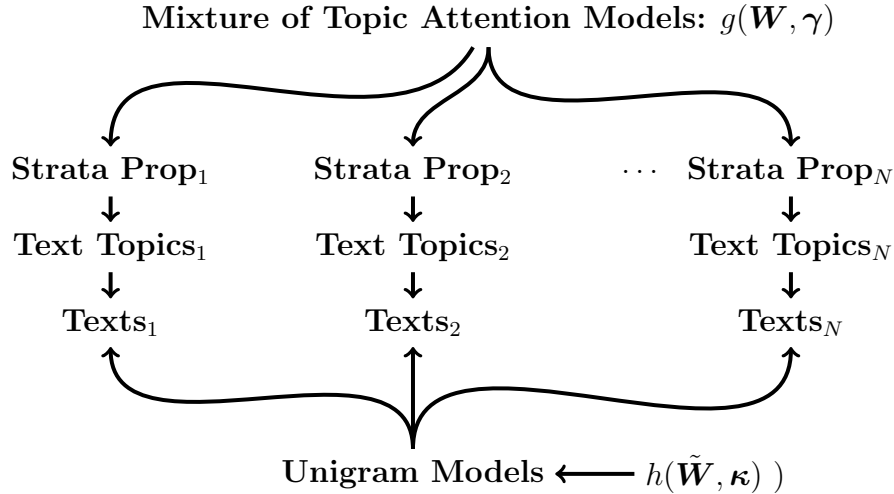


FIGURE 3.1: Illustration of the Structural Topic Model (Roberts, Stewart, Airoldi 2014)

with varying probabilities. The estimated probability here is, therefore, the probability that the document contains the topic, with the associated words that fall under that topic.

For this project, I will begin with a simple generative topic model designed to properly categorize the various documents that will be studied. I utilize a variation of the structural topic model (STM) introduced by Roberts, Stewart and Tingly (2014). This model classifies documents using a simple unsupervised model that accounts for variations in metadata about the documents. The varied inconsistencies within the documents I will be studying requires a careful control over the document-level covariates that this model would be able to control.

Given the complexity and the naturally occurring hierarchies within testimony data and in the text of bills, the structural topic model will be the best way of including as much information about strata covariates to make better predictions about the content of each document. Figure 3.1 draws out the logic of the structural topic model, and shows how it differs from the traditional LDA—the presence of the strata proportions allows for additional clustering along documents that share relevant covariates, and should not only improve model fit, but should also better approximate the process through which these words are selected to begin with.

Text-Reuse and Policy Ideation

After sorting the testimony and bill sections into clusters by the STM, I apply a text-reuse algorithm to determine what lines from testimony appear in the final language of Dodd-Frank. The text-reuse approach is a fairly new approach to the study of idea transmission in social science, but has been used to measure changes in statutes and regulations (Wilkerson et al. 2015; Haeder and Yackee 2015)

There are numerous machine learning algorithms that perform quite well in evaluating document similarity using “bag of words” approaches (Grimmer and Stewart 2013). However, machine learning scholars have found that text reuse, by using additional information about word or character sequence, is generally better at discerning more nuanced similarity. Instead of just looking at the number of shared words between two documents, word ordering matters.¹ Plagiarism software is the most common application (Hoad and Zobel 2003), but text reuse methods have found wide use in research and industry. It has been used in information retrieval to identify duplicate search queries; in communications research to study meme transmissions; in digital humanities research to trace the influence of notable books, and in music to compare reused sections of scores (Downie and Nelson 2000; Henzinger 2006; Leskovec, Backstrom, and Kleinberg 2009).

For this chapter, I utilize a versions of plagiarism detection software designed to be versatile enough so that it can be adapted for procedures beyond just detecting plagiarism in student work but also efficient enough that I can run the tens of thousands of bill section/testimonies combinations without slowing down indefinitely; WCopyfind (Bloomfield 2012). WCopyfind utilizes a simple string matching algorithm that allows for some built in flexibility (a leave-one-out provision to match near matches), all while catching each instance where a string shares in the same minimum five words consecutively. These shared sections,

¹ As an example, “bag-of-words” similarity measures would miss sentiments preceded by negations. There would be less of a difference between documents that argue for or against a given topic compared to ones that discuss orthogonal topics.

Table 3.1: An Example of Text-Reuse Between Testimony and Bill Language

Testimony House Hearing 110th Congress 45625	Bill SEC. 6013
<p>Specifically, I would emphasize the desirability of, first, standardization of ratings terminology so that all firms report ratings using identical terms. Second, industrywide standardization of stress conditions under which the ratings are evaluated. Thirdly, unambiguous quantitative correspondence between the ratings and the default probabilities under standardized conditions of economic stress. And fourth, standardization of terminology across asset classes so that, for example, a given rating applied to a municipal bond and a corporate bond will have the identical default probability under identical market stress conditions. One specific proposal that I would like to see investigated are what might be called ABZ ratings at a glance in which the three digits of a rating, instead of just being things that are made up, correspond to default probabilities under three different, well-defined levels of market stress. So the first digit, for example, could represent the default probability under normal market stress, the second digit under severe market stress and the third digit represents the default probability under extreme market stress. A 50 percent asset price drop, 10 percent increase in unemployment and so on.</p>	<p>An example to provide initial guidance for the study is a ratings symbol consisting of three digits, each of which corresponds to default probabilities under different levels of market stress as follows: (A) The first digit represents the default probability under “normal” market stress, characterized by normal economic fluctuations in addition to a 5 percent decline in asset value and 2 percent increase in unemployment. (B) The second digit represents the default probability under more severe market stress, characterized a 20 percent decline in asset value and 5 percent increase in unemployment. (C) The third digit represents the default probability under extreme market stress, characterized by a 50 percent decline in asset value and 10 percent increase in unemployment? standardizing credit ratings terminology, so that all credit rating agencies issue credit ratings using identical terms; (B) standardizing the market stress conditions under which ratings are evaluated; (C) requiring a quantitative correspondence between credit ratings and a range of default probabilities and loss expectations under standardized conditions of economic stress; and (D) standardizing credit rating terminology across asset classes, so that named ratings shall correspond to a standard range of default probabilities and expected losses independent of asset class and issuing entity.</p>
<p>Testimony of Bill Foster, PhD member of Joint Credit Reporting Taskforce.</p>	<p>Description of GAO report to be written underlying accountability of credit scoring criteria</p>

after removing some false-positives, make up the instances of idea reusing that forms the basis for the analysis in this chapter. The main conceit underlying this analysis is that shared language in such a clean way means at least a shared conceptualization of some piece of information. Extensive sifting to remove identical language that was either simple turns of phrase, agency names, or other obviously non idea centric language is what makes this type of analysis work.

Table 3.1 shows an example of a section of testimony text that has the same overlying idea content as the bill text, in this case, a discussion of how to rework credit scoring oversight to prevent abuses like there were in scoring the subprime mortgages.

Data

For this chapter, I decided to focus on one piece of very important legislation as something of a starting point: the Dodd-Frank Wall Street Reform and Consumer Protection Act. I take the series of hearings—61 total—from November 2009 through July 2010. Dodd-Frank works for this project as a first-pass for three reasons. First, the bill itself went through 5 different drafts, so comparisons can be made between drafts easily. Second, the breadth and depth of the testimony that occurred over the bill allows for sufficient variation in the content of what was discussed. The bill has a wide enough scope that the emergence of distinct topics should be readily apparent. The third justification for using this as the test case is that Dodd-Frank is an undeniably important piece of legislation, vital to the president’s first term agenda, and was the focal point for the reform of the major financial institutions and banks following the financial crisis of 2007-2008.

Dodd-Frank was designed to be the most ambitious reform of our major financial laws and regulations since the Great Depression and was written with the above mentioned bills explicitly in mind. Dodd-Frank, as laid out in President Obama’s “New Foundation” proposal, was supposed to bring order to the major regulatory bodies and generalize oversight in a more top-down approach, increase broad regulation of the financial markets (with particu-

lar note to look at derivatives), create a new consumer protection agency designed to provide a forum for adjudicating consumer and investor protection issues, increase the domain of the Federal Reserve and the Federal Deposit Insurance Corporation to deal with the winding down of bankrupt firms, and generally increase in the standards of regulation and oversight of credit rating and other independent auditing (Treasury 2009).

Dodd-Frank passed more or less along party lines, with a 223-202 vote in the House and a 59-39 vote in the Senate. What makes Dodd-Frank different than the other bills included here is that the time between the first draft of the bill—December 2009—to the final version—July 2010—was large enough that there were numerous major public hearings. Because this bill was a response to the crisis, and because it was exceedingly ambitious in its purview, the hearings provide a unique opportunity to study the role that outside testifiers have on the writing of legislation. During the seven month period the bill was being considered, there were 63 hearings on elements of Dodd-Frank, and, as such, every part was explored in extreme depth. The final result is a bill that fuses together the expertise of literally hundreds of speakers, all the while remaining a highly partisan and contentious bill. The partisan aspect to idea borrowing is immense with this bill.

For any topic modeling approach, we use the “bag-of-words” approach, meaning that the most efficient way to analyze document text is to look at the specific term frequencies and tendencies between documents independent of immediate context. Once documents are re-sorted into document-term-matrices, we strip out punctuation, numbers, and “stop words” (words that confer no additional information and are merely linguistic connectors or pronouns). After this, the words are all “stemmed,” taking the word to its most basic and shared root, essentially stripping part-of-speech suffixes.

For this iteration of the project, I fit separate structural topic models on the testimony surrounding Dodd-Frank and the drafts of the bills themselves. For the testimony, I treated each uninterrupted speech as a distinct document. In total, I had 12,000 unique documents, corresponding to 418 speakers (including Congressmen and Senators speaking at the

hearings). Additional information was collected about each speaker, including their basic demographics, their campaign contributions, and more specific information about their organizations. Table 3.2 shows the breakdown of the various speeches by the organization the speaker belongs to. Note the largest subgroups are the House and Senate, which makes sense given the testimony-response model of most hearings.

Table 3.2: Dodd-Frank Testimony Totals by Speaker Organization

	Number of Testimonies
Commodity Futures Trading Commission	135
Congressional Research Service	11
FDIC	295
Federal Housing Finance Agency	36
Federal Reserve	1171
Federal Reserve of Minneapolis	32
Federal Reserve of New York	13
Foreign Government	17
Government Accountability Office	66
House	2186
Local Government	254
Office of Thrift Supervision	24
SEC	285
Senate	2584
Treasury	1459
<hr/> Total Government	<hr/> 8568
Academia	629
Industry	2669
Interest Group	135
Think Tank	462
<hr/> Total Non-government	<hr/> 3895

For the five drafts of Dodd-Frank, I split the data up into each section for each version of the bill. The structural covariates included in the topic model estimation were the version of the bill and the sequence of the given section (operating off of the logic that the order in which sections are added to bills is non-random, and in fact corresponds to some general template for legislation). In table 3.3, we can see the distribution of the various sections by the version of the bill under discussion. Further, we can see the date the specific version was

introduced. Of note is the changing size of the bill: the version of the bill sent to the Senate from the House has the most sections with 518, whereas the final passage bill only has 268 sections, suggesting quite a bit of altering between the two moments.

Table 3.3: Dodd-Frank Section Totals by Version and Date

Version	Section total	Issue Date
Introduced in House	395	12/2/09
Enrolled in House	518	12/11/09
Referred to Senate	518	1/20/10
Amended in Senate	411	5/20/10
Final Bill	268	7/21/10

Because of the complexity of the data and the different document sizes, I fit two topic models of different pre-set topic levels (k) for the two sources. For the testimony data, I set $k = 40$, which is the same size used in other Congressional speech based research (Grimmer 2013; Grimmer and Stewart 2013). For the actual bill model, I set $k = 30$, which corresponds better to its smaller size. I used a variety of different k 's for both models before settling on the ones chosen for this chapter: evaluated using the standard model based characteristics described in Stewart, Roberts, and Tingly (2014). Sample fit statistics—including semantic coherence, exclusivity, and held-out-likelihood—are discussed in an online appendix, and are compared to various other potential settings for k there. I believe that the k 's set for the two models generate the most coherent models that also fit the various evaluative criteria laid out.

After fitting the topic models, I combine the two data sources into a single master corpus. From that I fit a structural topic model that incorporates information from both testimony and bill sections, which allows for comparability between documents. To deal with the different linguistic tendencies of the two sources, I utilize a content covariate that separates the two; this allows for the model to learn which words are unique to either testimony or bill text and generate topics based only on words that appear in both. I then run the testimonies through WCopyfind to compare with the bill sections. From that, I count not only how many

bill sections a given testimony is connected to but also how many phrases they share.

Results

To begin, I limited the analysis to a handful of topics for each model, to show the various ways the information here can be explored. After discussing a few interesting topics for both the bills and the testimony, I present the results of modeling the text-reuse algorithm’s output. I combine this with other information collected about the speakers to determine what predicts highly “influential” testimonies and what that reveals about the drafting of Dodd-Frank.

Test-Reuse Results

To start to understand what is going on during the hearings and the concomitant drafting of Dodd Frank, I fit the structural topic model on both, incorporating information from these pre-existing topic models I ran on both separately.² This model was then used to classify the documents by the policy region being discussed. Figure 3.2 details the topics for this model. After running a series of model fit diagnostics, I settled on a topic model with 60 topics.

Notably for this model, I included the “content covariate”, to distinguish between bill sections and testimony texts. The logic of the content covariate is that it is possible to get more coherent and more stable topics if we can capture variance in the ways in which a given topic is discussed (Roberts et al 2014). Since the language of testimony is going to serve a different function than the language of legislation, we can assume that the ways in which the topics are used will vary, a problem solved by fitting a content covariate in the model. I also include a time component to the model, because the general agenda of Congress is likely to have shifted somewhat during the months Dodd-Frank was considered.

² The results are available upon request. These separate models, which are not comparable because of the logic of sampling from the simplex that underscores the distributive properties of topic modeling, were used to gain insights into what was being discussed in both, and how they might intersect. Since these models were run on each text source separately, modeling for more particular covariates could be done without threatening convergence and stability. The joint model, reported on here, ended up being remarkably stable. Diagnostics for this model are included in the appendix.

I ran WCopyfind on testimonies and the bill sections, manually cleaning the matches identified by the algorithm, to eliminate false positives. Below are the results of counting the number of bill sections to which a given testimony is connected.

To best understand what predicts a testimony sharing content with a bill section, I ran distinct models, all on the same variables: 1) a Logit model on whether or not a given piece of testimony has any connections at all, 2) a negative binomial model on how many bill sections a given testimony is connected to, and 3) two other negative binomial models on how many overall connections a given piece of testimony has (with slightly different specifications).

One main hypothesis of this chapter is: If speakers are highly trustworthy, not through partisan signaling or by purchasing access but through academic expertise, then MC's will borrow their text most closely. To best test whose influencing the final language of the bill, I included several covariates related to the speakers themselves, which should give us an idea as to who is influential. The variables I included were all dummy variables: 1) if their primary industry was business (including bankers and people working for business related lobbying groups), 2) if they had made a campaign contribution during the 2008 or 2010 elections, as an admittedly weak proxy for if they were involved in trying to buy a seat at the table, 3) if the speaker was primarily employed in academia, as a professor (usually in either economics, business, or law), and 4) whether or nor the speaker lived primarily in the DC metro area (abbreviated here as DMV for DC, Maryland, and Virginia).

The second main hypothesis is that if the topic the speaker discusses is highly complex and urgent, then MC's will also borrow the text most closely. The other two variables included here are combinations of several topics that are policy relevant, and vary on the level of complexity the topic suggests, which I've termed "High Complex" and "Low Complex" respectively. "High Complex" is a combination of Topic 2 (risk, redund, system, profil, manag), Topic 31 (enforc, disproportion, agenc, complianc, necess), and Topic 53 (deposit, lien, leas, communiti, reorgan). "Low Complex" is a combination of Topic 19 (freedom, fine, print, anyon, virtu), Topic 14 (effort, forum, ethnic, demograph, acorn), and Topic 13

Top Topics

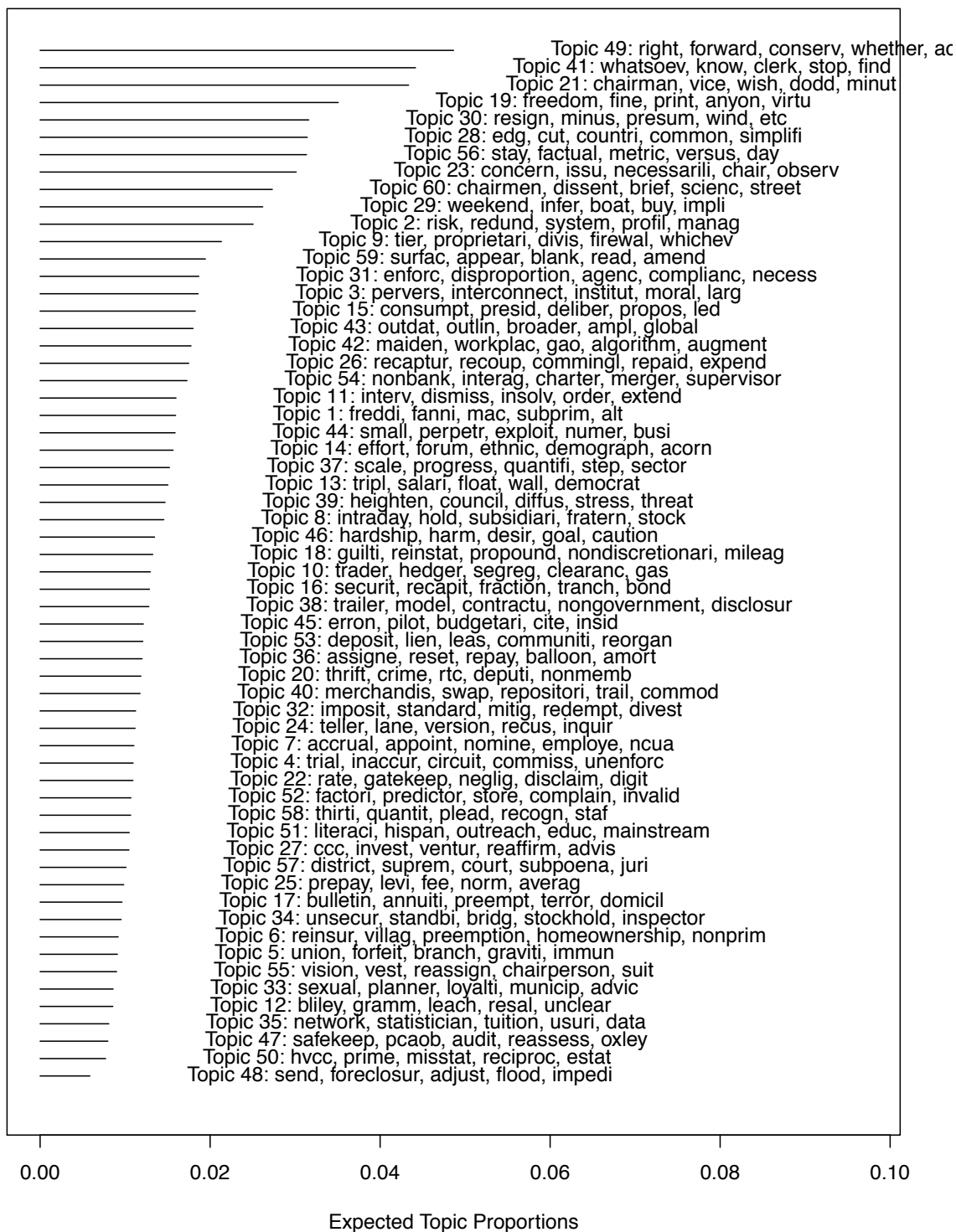


FIGURE 3.2: Topics for Combined STM

(tripl, salari, float, wall, democrat). One can justify using these sets of policy topics as representative of highly complex topics versus very low complexity topics, but the biggest evidence to the effect is to look at the bill sections dominated by these topics. High complex topic heavy bill sections tend to be longer and have more arcane language than low complex ones. As a validation approach, principal components analysis was also performed on the posterior probability matrix of the topics. The first dimension (which had high loadings on all the high complex topics and negative loadings on many of the low-complex ones) was used in lieu of these categories in model 4 (and denoted with ‘pcomps’).³

As seen in Table 3.4, results of these models are clear in several regards. First of all, the effect of living in the DC area, and of being in academia is always or mostly statistically significant, suggesting something about whom is able to get the attention of Congressmen. The academic effect here may be partially due to the complexity of what they are discussing in hearings as compared to occupations (e.g. a lobbyist), and the lack of personal stake makes their own expertise more credible. We will return to the complexity issue shortly. The DC area variable being consistently statistically significant probably relates to the overall efficacy of interpersonal interaction beyond just the confines of campaign contributions and lobbying; it suggests that a more network-based form of interpersonal connectedness is important (see Chapter 2). Or it could be that my other proxies for this legislative capture/lobbying effect (e.g.,donations) are too crude to be useful here and requires additional refinement.

The other interesting finding was the relationship between the topic probabilities in the testimony and the levels that their text were borrowed in bill sections. Consistent with predictions, we find diametrically opposed results in totals and counts between high and low complexity topics. The probability that a given document is going to share text and the estimated number of bill sections and total segments shared with that document is

³ The set of complex topics hardly unique, and indeed, the model has been run with many combinations of these topics plus others. The results did not vary when different credible combinations were tried. The topics presented here represent the clearest theoretically joined sets. As to the principal components analysis, more details about the dimensions are available upon request.

Table 3.4: Text Reuse Modeling

	<i>Dependent variable:</i>			
	reuse <i>logistic</i>	count <i>Negative Binomial</i>	total <i>Negative Binomial</i>	
	(1)	(2)	(3)	(4)
high_complex	4.172*** (0.830)	5.940*** (1.768)	7.858*** (2.336)	
low_complex	-7.716*** (2.121)	-12.852** (5.184)	-9.582* (4.890)	
pcomps				11.067*** (3.070)
Donate	0.095 (0.112)	0.043 (0.238)	0.113 (0.272)	0.405 (0.337)
Academia	0.513*** (0.182)	0.831** (0.398)	0.837* (0.496)	0.956 (0.602)
DMV	0.084 (0.110)	0.751*** (0.242)	0.939*** (0.269)	1.033*** (0.330)
Constant	-2.520*** (0.206)	0.285 (0.477)	2.288*** (0.504)	1.827*** (0.346)
Observations	5,613	5,613	5,613	5,613
Log Likelihood	-1,395.687	-16,154.170	-57,701.100	-55,923.180
Akaike Inf. Crit.	2,803.374	32,320.330	115,414.200	111,856.400

Note:

*p<0.1; **p<0.05; ***p<0.01

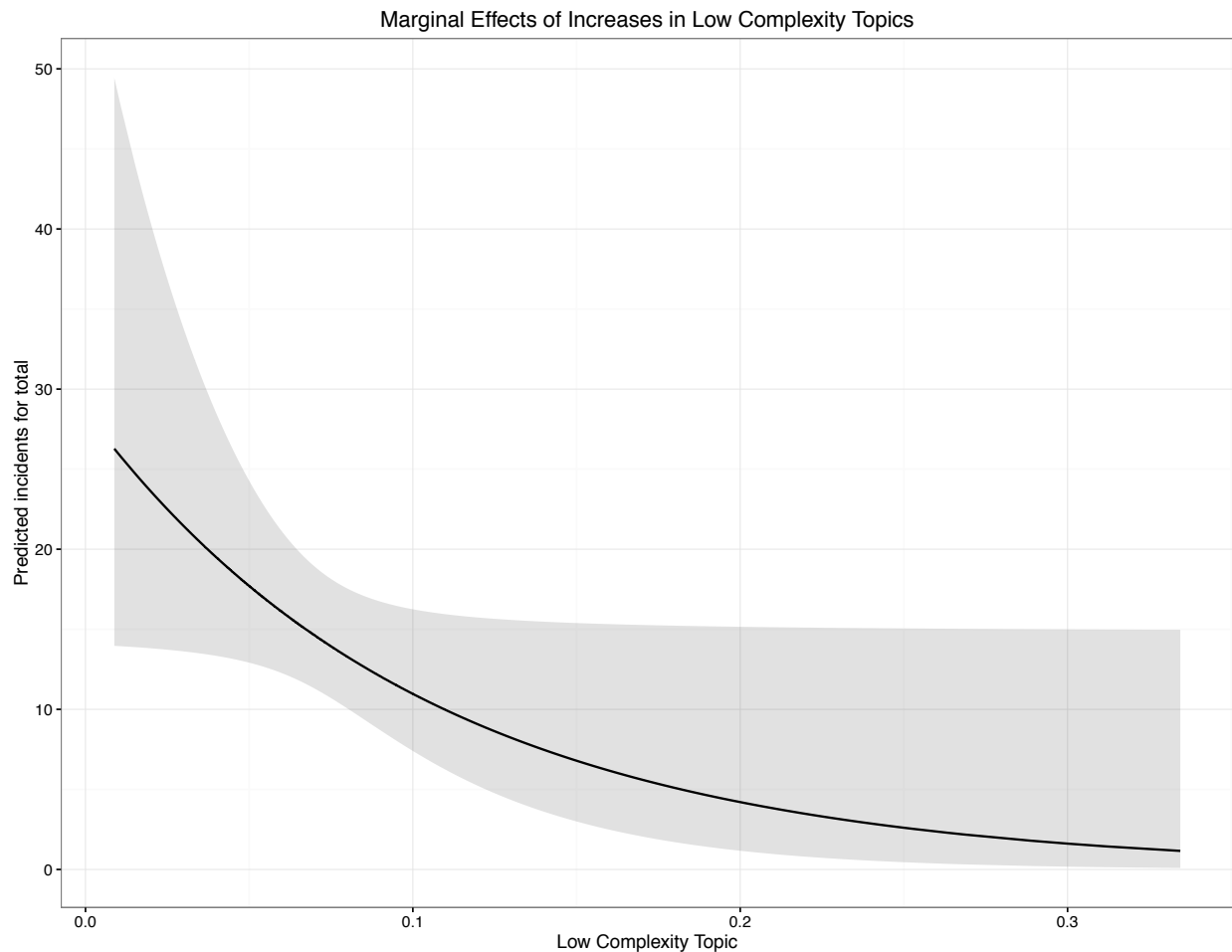


FIGURE 3.3: Predicted Counts by Topic group: Low Complexity Topics

highly related to the probability that the testimony discussed is a complex topic. An inverse relationship is found for low complexity topics. The more likely a testimony is to talk about this given topic, the less frequently it will share text with bill sections, and the fewer sections of text overall will be shared with that testimony in general.

Figures 3.3 and 3.4 explore this in greater depth. Looking at the predicted counts of shared text a given document will have in Figure 3.3, the total drops dramatically as the document is more likely to contain low-complexity policy topics. We find, again, a nearly opposite result when looking at their probability to discuss highly complex topics. The total estimated shared texts increases exponentially. This may in part be due to the general complexity of Congressional action regarding these two very different tasks. With reforming

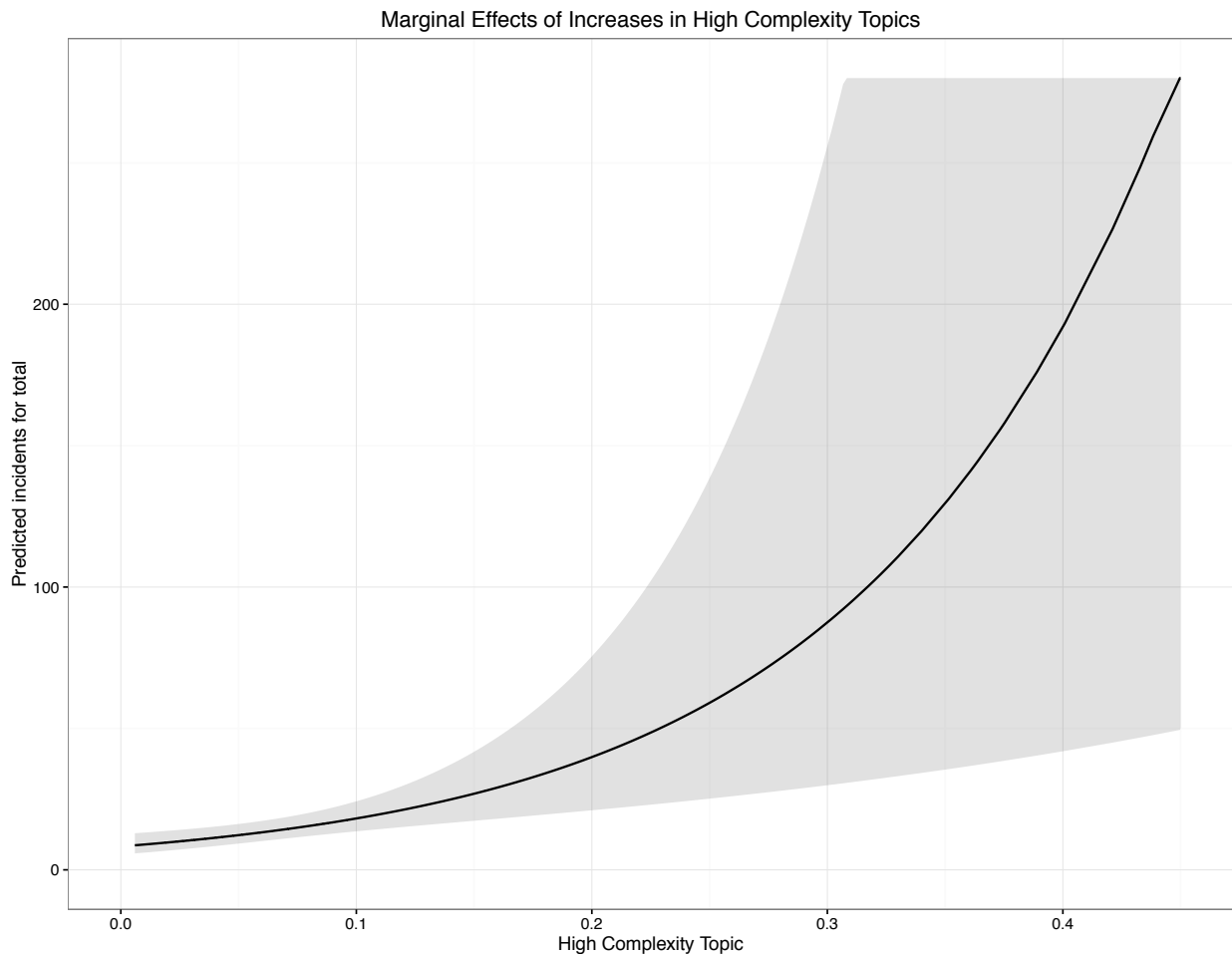


FIGURE 3.4: Predicted Counts by Topic group: High Complexity Topics

the Federal Reserve, the complexity of the issues going into that decision are enormous, and the ramifications of making a serious mistake are severe and could indeed plunge our economy into a depression. Because the institution being discussed here is both highly complex and potentially the most consequential, Congress, when writing legislation designed to affect the Federal Reserve, defers to the expertise of the people called to testify.

Again, as a validation strategy, we look at a dimension extracted from the principal components analysis discussed above. Figure 3.5 depicts the result of this predicted counts model. We see a very similar shape to the distribution as we had in the high-complex topic (even the fatter tail as the overall probability increases). The strength of this result suggests that the underlying dimension seen above is able to be replicated through a more

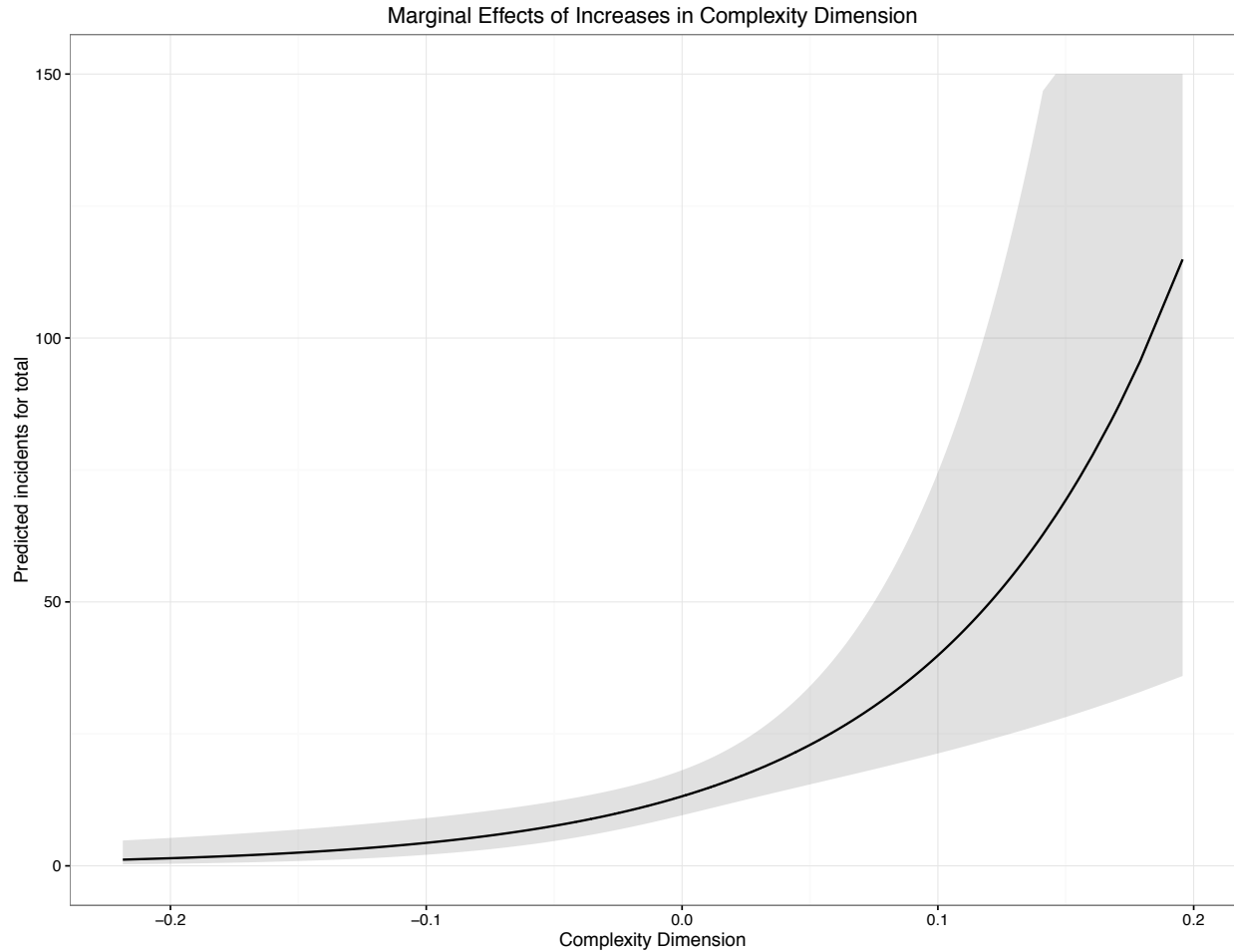


FIGURE 3.5: Predicted Counts by Principal Component of Topics

assumption-free approach to modeling language. The higher complexity topics, by virtue of dealing with complex financial regulations and instruments, force MCs to think outside of their own personal knowledge base and seek out someone with the necessary expertise to guide them on these issue. Consider the example from Figure 3.1 again: restructuring the credit scoring mechanisms requires far more knowledge about how these scores come about, what the impact of minor changes will be, and so forth, than any MC would have. The results here simply reinforce this casual observation and suggest future tests for getting at more precise mechanisms of idea borrowing in legislation.

Conclusion and Next Steps

This chapter should reinforce the importance that the substance of bills—namely the policy ideas written into law—should play in our discussions of what Congress does. In this chapter, I have demonstrated the ways in which Congress engages in idea borrowing and how that impacts the writing of law. By borrowing ideas more forcefully from testimony dealing with complex topics, Congress shifts the focus to the experts testifying, an important act of implicit delegation that suggests Congress knows its own limitations better than it otherwise might appear. This is consistent with the idea that this delegation is done strategically, and dovetails nicely with discussions of how this occurs in the bureaucracy literature (McNollgast 1987; McNollgast 1989; Epstein and O’Halloran 1999; Huber and Shipan 2002).

The net effect presented should be more on the potential for future research than the conclusions drawn. Topic modeling provides a solid framework to evaluate an interaction between the language of testimony and final bills, but it is one that requires a very precise and careful research design to implement fully. This project gives an early perspective on mapping these two very diverse sources of textual data and should lead to a fruitful research agenda in the near future.

The clearest point for extending this chapter is to better capture what it means to have a text-reuse connection and to use that information to better test the theories presented in the beginning of this chapter. The first major step will be incorporating additional outside sources of text for the non-government testifiers, to lay out the pool of ideas from which they are sampling from when they testify. Since the ideas discussed in testimony are not drawn *ex-nihilo*, we must try to trace the ideas even further removed. This allows for a better mapping of the universe of policy ideas. Once we have the ideas properly identified, we can better understand the universe from which they are pulled, as we map their transmission from outsiders to MCs in a singular network. Taking the ideas as information metaphor to its logical conclusion would necessitate a network centric vision of what idea transmission

looks like, and this should provide the basis for such a study.

Exerting this further beyond this project, this text-reuse method could be used to trace ideas in many different circumstances and could be implemented in a more “big data” related way. While Dodd-Frank is certainly an interesting piece of legislation to test this on, it is only a single case study and has all the limitations that such a study would have in its exculpatory powers beyond the single case. It would be worthwhile to see this same framework applied to other bills of varying salience and relevance, also to compare this to other important bills under a different legislative framework (based on committee assignments, divided government, divided houses, etc.). Doing a more systematic study of where ideas in legislation come from has the potential to seriously challenge a lot of the assumptions we make about the legislative process and, if this can be streamlined efficiently, could create seriously augment dominant theories of what Congress does and who is making the crucial decisions. Future studies of Congressional behavior will have to take into account what is being done within the bills that are being passed, and more systematic studies of the act of writing legislation will only improve our understanding of this.

Money Doesn't Talk (It Swears): Tracing Ideation in Major Financial Legislation

Introduction

It is not hard to argue that ideas matter; indeed, it probably goes without saying that ideas are inseparable from our conceptualization of politics. Even in the most material driven account of political decision-making, some role for ideas remains absolutely inescapable. Understanding how and when ideas matter, and when they intrude upon the complex procedure that is policymaking, is crucial to understanding how our political system functions. Being both shaped by and shaping the ideas undergirding major policies creates an abstruse feedback loop that is hard to penetrate. Even more important is to understand the impact that ideas have how and when do ideas have consequences. Thi

It is fair to say that, from a policy perspective, very little in American politics has had a greater consequences on both our politics writ large and on the day-to-day lives of our citizens than the Great Recession of 2007-2008. While the significance of the recession is easy to narrow down (since it is so comprehensive and widespread), the causes of it are markedly more difficult to pinpoint. Indeed, serious analyses of the crisis have sometimes pinpointed

the exact opposite policies as causes of the crisis,¹ making narrowing down a particular culprit uniquely difficult. Since answering this essential policy question is probably outside the purview of the author, I focus on something far more tenable,² namely where did the ideas underlying the financial crisis come from? And how are matters being handled differently now?

In this chapter, I use a method for tracing idea flow introduced in Chapter 3, and implement it over four key pieces of financial legislation, two laws mentioned in particular as playing a key role in the shaping the financial crisis, and two laws that act as a direct response to the crisis. I implement a variation of the topic modeling algorithm for generalizing topical trends regarding texts for the testimony for each law. Once these texts are classified by a simple clustering method on the posterior probability matrix generated by the topic model, I use a text reuse algorithm (adapted from a popular form of plagiarism detection software) to track instances where identical text appears in a section of a bill that first appeared in testimony (Bloomfield 2012).

This approach addresses two key points: 1) what does the legislative landscape look like during the writing of complex financial legislation, and 2) under what conditions do members of Congress (MCs) borrow ideas when writing legislation, all the while analyzing legislation that remains key understanding the financial crisis. By studying these four laws—Gramm-Leech-Bliley, Sarbanes-Oxley, Troubled Asset Relief Program, and Dodd-Frank—I am able to hold the general subjects of the laws constant, since they are all major financial legislation, while exploring conditional differences. Indeed, the major players remain the same between bills, but differences in how ideas are reused show consistency to the point where stronger conclusions may be drawn.

¹ See in particular the discrepancy over the regulation of the holding allowances by commercial banks, where notable economists have offered widely contrary accounts as to its impacts (Posner 2009; Krugman 2012 contra Meltzer 2009). An overview of the major disagreements over this particular point can be found in Crawford (2011).

² And, given the authors background, far more political

Background

Ideas in Politics

My approach to discussing ideas in this chapter is directly related to the concepts outlined in Chapter 3. This conceptualization takes ideas as an important type of information and, instead of acting outside realm of material interests, is instead part and parcel to rational behavior. An idea is a subset of information—it requires information to take hold—but is itself more than just information. In this approach, I treat an idea as a mental representation of a piece of information, which is inherently personal but known observationally. It is different than information in that it requires human interaction; a piece of information is simply a fact (or a series of facts), but an idea is a reflection of how that relates to something more abstract.

This is similar to the notion of “conceptual blending”, which argues that the only way we understand abstract concepts is by combining concrete concepts we understand better; I argue that ideas are a realization of the conceptual blending (Turner 2001; Mackay 2003; Fauconnier and Turner 2008; Turner 2014). Conceptual blending builds off of the notion that all forms of higher level human cognition are derived from representational language, and are based on the implementation of practical metaphors. Consider how we think about what an “ideal point” is, to choose perhaps the most discussed construct in the American politics literature. We use the spatial metaphor to illustrate that a persons ideology exists relative to another person; we ascribe notions like directionality and dimension to this as if these are meaningful constructs outside of the spatial metaphor. Even the very idea of a “left wing” and a “right wing” pull from multiple metaphors; we blend together the ideas of spatial dimensions to politics with historical language about physical wings to represent location. By combining two visuals we understand already—left-right spatial dimension and physical wings in a hall—we create a meaningful intellectual representational language that delineates a series of ideological positions; we create the idea of a left and right wing ideology.

For this chapter, I argue that ideas absolutely must be transmitted like information. As such, we can use methods that look at specific segments of text as the basis for ideas. Indeed, the reuse of specific words or phrases can indicate that the speaker is sharing in all of the necessary meta-context and thought behind the ideas in question. Thinking of how ideas are transmitted as being parts of an information transfer process allows us to model it using frameworks that are robust to complex interactions, including formal representations like hearings. Information theory gives us the idea that this transmission is itself reproducible in a formal setting.

The Role of Congressional Hearings

The main conceit of this chapter is that we can compare language used in hearings to language in bills and make reasonable inferences from that. For this to be true, the role that hearings play must be informative, not strictly demonstrative, and the language from these hearings must not just be “cheap talk”, but instead play a substantive role in the hearings process. Much analysis has been done operating off of that exact assumption, that hearings are informative, and this would not put my efforts outside the mainstream in political science (Diermeier and Feddersen 2000; Esterling 2004; Esterling 2007 as examples). Below is a discussion justifying the use of Congressional hearings as a source of important information regarding the writing of legislation.³

Hearings are informative because the committees wanted it that way. Committees carefully orchestrate hearings so that the record reflects their work on legislation and so that the hearings shape legislators’ understanding of the issue in a manner the committee members find most useful; it isn’t difficult, given this, to argue that any hearing is informative if the committee wants it to be (Esterling 2004). Since the chair of a committee has complete calendar powers over the hearings and controls the exact timing of all witnesses, while committee staffers go to great lengths to prepare witnesses to testify, committees then are

³ Paraphrasing and adding to a very thorough discussion of this very issue in Esterling (2004).

incentivized to use the hearing demonstratively. As Esterling (2004, p. 252) puts it:

“Hearings show that legislators have developed the necessary expertise to write good policy, perhaps to expand or maintain committee jurisdiction (Talbert, Jones, and Baumgartner 1995), to extract concessions from the legislature as a whole (Krehbiel 1992), or to help with reelection efforts (Bianco 1994)... Hearing[s] reflect the committee’s relative need for information on an issue. That is, hearings will be either informative or not informative depending on the committee’s special interests regarding and informational needs for the policy topic. Precisely because committees prepare very carefully to ensure that hearings meets goals, the hearings are a valid data source for gauging the nature of committees’ interest in policies.”

Given the highly contentious nature of the policies being proposed in the financial legislation hearings under discuss in this chapter, it is nearly impossible to think of them as anything but having high ”informational needs.” These implications are essential to building a theory of idea transmission.

This account runs contrary to the traditional view that sees hearings only as a means for advocating politically convenient policies in a demonstrably public way, i.e., that it is only an exercise in position taking or advertising (Mayhew 1974). This view of committees sees the role of the hearing as just assembling a list of witnesses to advocate the chair’s prior position. By this logic, committees care only about the positions taken during the hearings, not the veracity or the validity of the positions being taken. This presupposes something dangerous about the nature of lawmaking, an argument that any policy effects are incidental to the writing and passage of policy (which would flow from following Mayhew’s logic in its strictest sense). If one expands this definition to include *any* preferences for writing “good policy” then the act of having the hearings must perform some information transmitting task and should select for quality information (Fenno 1973; Esterling 2004, see Diermeier and Feddersen 2000 for discussion specific to hearings). One can argue that committee hearings simply reflect what committees already know or that committee members ignore

what is said in hearings, but this cannot be true for hearings over issues that were yet unknown, nor for things that are traditionally outside the purview of the normal MC.

Again, though, as to the relevance of hearings as more than just “cheap talk”, but as a forum for meaningful exchange of information, consider what the members themselves argue. As Kingdon states, “Congressmen themselves believe the hearings do provide them with the bulk of information needed to legislate, and the hearings do affect their decisions” (1989, p. 213). Furthermore, MCs behave as if their fellow members who have sat through and participated in the hearings are themselves more knowledgeable about an issue than they otherwise would have been. Again, referring to Kingdon, he finds that members gain influence over their colleagues through their participation in hearings: “Congressmen often refer to the opportunity that their colleagues on the committee had to attend the hearings and listen to the experts” (1989, p. 85-86). DeGregorio, regarding committee staff members, states that “According to a sizeable proportion of senior staffers, legislators may decide whether or not to pursue a policy on the strength of the arguments that they receive in testimony” (1992, 978-79). Given what we know about the increasingly limited time and resources at the disposal of the committee staff, it is implausible at best to assume that committees already have full information about what will be said at hearings, as the staff and MCs will often lack both the specific information presented and the expertise to adjudicate directly regarding the quality of expert testimony before the hearings proceed (Esterling 2004).

Given the above, it is clear that ideas are highly salient in politics, that Congressional hearings are platforms for meaningful exchange of ideas, (particularly when the topic is fraught with complexities and has wide ranging consequences), and that the landscape of financial legislation over the past 20 years qualifies as such. With all that, it becomes necessary to raise the questions: under what conditions would MC’s use ideas from experts at these hearings to actually write up legislation, how is this done, and what are the consequences?

Theory

The main argument in this chapter is that idea borrowing occurs in Congress when there is a purposeful agreement on the substance of a bill but potential confusion over either implementation or just in the complex nature of the bill itself. This argument suggests that the subject being deliberated drives how MCs interact with the ideas presented and that the most complex sets of ideas must come from external forces, barring unique Congressional expertise (which, as Esterling 2007 points out, is rather rare even in the more conceptually straightforward policy topics and we can assume is surely undercounted in the high levels of abstraction required to deliberate over financial legislation).

To get to this point, a few key premises must hold. First, it is assumed that MCs see no difference in where the text of a bill comes from, as long as it performs the task required at hand. This requires a bit of unpacking to get to, but it is essential to understanding how MCs use ideas presented during hearings in the writing of bills. More fundamentally, this holds true because MCs are driven to maximize both the desirable outcomes and to minimize the amount of work needed to be done to create an idea anew. Minimizing the amount of work needed to be done would not be a terribly controversial point, but it does build upon the fact of decreasing staff sizes in Congress combined with increased time demands for campaigning, any relief of the Congressional workload would be expected to be sought out immediately. Given this, it needs to be emphasized that MCs are weary of reinventing the wheel; if something from a previous bill does what they want it to do, they will reuse it (Wilkerson 2015).

Given these premises, it follows that MCs are willing to shirk the responsibility of writing specifics if they have experts' direct wording they can pass off as their own. The conditions for this are relatively clear: the experts either represent a vested interested the MC is trying to properly ameliorate (the focus of Esterling 2004), or the interest group landscape is either sufficiently barren (unlikely in the case of financial legislation) or overly fraught

with competing interests to make any clear signal as to the politically expedient behavior less than clear. Put it another way, when the signals are so abundant because the issues represented are so complex, MCs may receive too many to make out clearly and will be receiving jammed signals. Disentangling them specifically would require more effort than just listening to the hearing testimony and instead adjudicating the case clearly.

Under what conditions will they borrow ideas from external experts most readily then? It follows that it will come first when there are no old ideas to borrow from first (specific policy implementation hasn't been done yet). This suggests borrowing as shown in Wilkerson (2015) is more difficult. Idea reuse will also occur when the issue is of significant salience and complexity such that doing a poor job is both undesirable and difficult to avoid. This follows from having meaningful policy preferences above. If the consequences (at least as far as the MC is concerned) of not achieving desirable policy outcomes are severe, *and* the issue falls under the realm where known experts will have greater ease in solving the problems and have a better grasp of the policies than the MCs, then and only then will MCs find it necessary and efficacious to borrow language from the experts. Policy complexity is conceptualized similarly to how Elgie and McMennamin (2005) discussed it, insofar as it refers to policies that require expertise to understand and implement, or, as they summarize it, "the greater the issue complexity, the greater the need for policy experts" (Elgie and McMennamin 2005 pg. 545). While the theory outlined in this chapter does not rely on a reductionist definition of policy complexity, it does find evidence of exactly that relationship.

Linking Legislation and Ideas

With these situations clearly laid out, there are some implications for the linkages between legislation and ideas that need to be made before a general theory about ideas can be properly claimed.

The process that legislation goes through before testimony can begin is rather straightforward. A bill is proposed, presumably with ideas present in the text. This bill is then

assigned to a specific committee or committees, where it is either rejected or assigned relevant subcommittees. The heads of the committees and subcommittees then call in experts to testify in their respective hearings in order to perform any one (or more) of three tasks. First, they address specific concerns as representative of an industry or constituency who desire Congressional action in order to ameliorate any concerns. Second, they speak to expertise on a given issue that would help in the writing of important legislation. Third, they are called to testify as part of Congressional oversight, and the testimony is partially a public excoriation or pseudo-trial (McCubbins and Schwartz 1984; Hall and Wayman 1991; Epstein and O'Halloran 1995; Balla and Wright 2001). These motivations determine the landscape of those who are called to testify, and the time demands of hearings in Congress, which are the determinates of the legislative agenda.

Using the concept of a “policy idea” adapted from Adler and Wilkerson (2012), I argue that it is through the selective choice of expert testimony made by committee and subcommittee chairs that determine what policy ideas make it into the domain of legislative discourse. They either use it to reinforce ideas already present in early drafts of bills or to introduce a potential new set of ideas to make the legislation better fit their agenda. Since bills can be thought of as “vehicles” for policy ideas (Adler and Wilkerson 2012) and that “many if not most important bills are largely drafted by committee staff under the supervision of the committee or subcommittee chair” (Evans 1991, Kaiser 2013, Wilkerson, Smith and Stramp 2015), it is reasonable to extend this to say that MCs define the parameters of what policy ideas to sample from by their choices of who to call to testify. Broadly speaking, legislation is a product of ideas transmitted through the strategic engagement of policy experts through committee hearings and testimony.

The nature of transmitting ideas from people testifying requires a model for understanding how MCs are able to adjudicate which information from people testifying is reliable and will lead to achieving the public policy goals they desire. This involves a hierarchical process, in which those who are called testify are not in fact selected randomly but are systematically

selected by the leadership of the subcommittees covering the issue space in question.

Based on this, we are able to derive some correlational expectations that we would expect to be met in predicting which speeches will contain text that is later reused in the writing of a bill. The topics that will induce the most text reuse will be those that deal with topics that are inherently complex and demanding of high level of attention, or how certain elements of the law will be implemented, such as regulatory structure dealing with a particular industry. Beyond those limited parameters, idea reuse should be relatively rare; as a part of how bills are being written, actually working in exact words and phrases used by witnesses would only occur when it has to, and those situations seem the most clear indications that it must. If my theory is correct, then the actual rates of reuse for a given speech will vary considerably by how complex the topic discussed is and how much the particular speech deals with implementation and regulation.

Methodological Approach

To properly analyze idea tracing and the topic of a given testimony, I utilize two distinct methods from the text-as-data toolkit: text reuse classifying for the idea tracing, and the structural topic model for the understanding the content of a given testimony. To sort the testimony and bill sections into meaningful categories (to insure that the text-reuse is not just picking up non-substantive language), I use a topic modeling algorithm, fit to testimony and bills jointly, then apply a text-reuse approach to find exact textual overlap between testimony and bill sections.

Methods: Structural Topic Model

Probabilistic topic models are a collection of algorithms that aim to discover the hidden thematic structure in large archives of documents. Topic modeling is a fundamentally Bayesian approach to discovering the underlying logic and structure of documents by analyzing the words of the original texts, which in turn is used to determine the themes that run through

them, how those themes are connected to each other, and how they change over time. Topic modeling relies on detecting the hidden structure in texts, in order to bring out the ideas behind them, while assigning posterior probabilities that any given document contains any generated topic (Blei, Ng, and Jordan 2003; Blei and Lafferty 2007; Blei 2012).

Topic modeling is a way of automatically discovering topics that these documents contain. For example, suppose we want to find two topics, A and B. To do this, a topic-modeling algorithm assumes that the author of the documents set out to write these document with “ n ” words (with the words being distributed by a Poisson Distribution) that contains a set topic mixture of the two different topics (distributed as a Dirichlet Distribution), and then using a multinomial distribution to pick words within each topic. The model then reassesses from the documents to find the underlying set of topics the author wanted to convey and does so with varying probabilities. The estimated probability is then the probability that the document contains the topic, with the associated words that fall under that topic. The virtue of the mixed membership characteristic of the topic models I will be using generally is that it allows for non-restrictive “blending” of topics which more naturally reflects how we utilize language.

For this Chapter, I will begin with a simple generative topic model designed to properly categorize the various documents that will be studied. My inclination is to use a variation of the structural topic model (STM) introduced by Roberts, Stewart and Tingly (2014). This model classifies documents using a simple unsupervised model that accounts for variations in metadata about the documents. The varied inconsistencies within the documents to be studied requires a careful control over the document-level covariates that this model would be able to control.

Given the complexity and the naturally occurring hierarchies within testimony data and in the text of bills, the structural topic model will be the best way of including as much information about strata covariates to make better predictions about the content of each document. Figure 3.1, which is on page 39, draws out the logic of the structural topic model

and shows how it differs from the traditional LDA—the presence of the strata proportions allows for additional clustering along documents that share relevant covariates and should both improve model fit and better approximate the process through which these words are selected to begin with.

Text-Reuse

After sorting the testimony and bill sections into clusters by the STM, I apply a text-reuse algorithm to determine what lines from testimony appear in the final language of each of the pieces of legislation. The text-reuse approach is a fairly new approach to the study of idea transmission in social science but has already been used to measure changes in statutes and regulations (Wilkerson et al. 2015; Haeder and Yackee 2015).

There are numerous machine learning algorithms that perform quite well in evaluating document similarity using “bag-of-words” approaches (Grimmer and Stewart 2013). However, machine learning scholars have found that text reuse, by using additional information about word or character sequence, is generally better at discerning more nuanced similarity. Instead of merely looking at the number of shared words between two documents, word ordering matters.⁴ Plagiarism software is the most common application (Hoad and Zobel 2003), but text reuse methods have found wide use in research and industry, having been used in information retrieval, to identify duplicate search queries; in communications research to study meme transmissions; in digital humanities research to trace the influence of notable books, and even in music to compare reused sections of musical scores (Downie and Nelson 2000; Henzinger 2006; Leskovec, Backstrom, and Kleinberg 2009).

For this chapter, I utilize a versions of plagiarism detection software, WCopyfind, that is designed to be versatile enough that it can be adapted for procedures beyond just detecting plagiarism in student work but also efficient enough that I can run the tens of thousands

⁴ As an example, “bag-of-words” similarity measures would miss sentiments preceded by negations. There would be less of a difference between documents that argue for or against a given topic than ones that discuss orthogonal topics

of bill section/testimonies combinations in a reasonable amount of time (Bloomfield 2012). WCopyfind utilizes a simple string matching algorithm that allows for some built in flexibility (a leave-one-out provision to match near matches), all while catching each instance where a string shares in the same minimum five words consecutively. These shared sections, after removing some false-positives, make up the instances of idea reusing that forms the basis for the analysis in this chapter.⁵ The main conceit underlying this analysis is that shared language in such a clean way means at least a shared conceptualization of some piece of information. Extensive sifting to remove identical language that was either simple turns of phrase, agency names, or other obviously non idea centric language is what makes this type of analysis work.

Table 3.1, on page 41, shows an example of a section of testimony text that has the same overlying idea content as the bill text: in this case, a discussion of how to rework credit scoring oversight to prevent abuses like there were in scoring the subprime mortgages.

Data

For Chapter 4, I focus on four pieces of legislation that are absolute necessary to understanding the lead up and response to the financial crisis and recession of 2007-2008. The four bills selected were the Gramm-Leach-Bliley Act (also known as the “Financial Modernization Act of 1999”), Sarbanes-Oxley (also known as the “Public Company Accounting Reform and Investor Protection Act” in the Senate and the “Corporate and Auditing Accountability and Responsibility Act” in the House), Troubled Asset Relief Program Act (also known as the Emergency Economic Stabilization Act of 2008), and the Dodd-Frank Wall Street Reform and Consumer Protection Act. The first two laws were passed well before the financial crisis

⁵ The removal of false positives is one of the more important approaches to making sure this method works. The single largest source of false positives came in naming either longer agency names or older legislation, so filtering through these wasn’t terribly difficult. Doing so for each piece of legislation was a fairly time consuming task, but after removing all strings that were just agency names, the false positive rate dropped dramatically. An actual idea transmitted would include something performing a specific political task, changing the authority of a given agency, or making a specific policy recommendation.

and have been mentioned as potential causes of the financial crisis; the second two represent two main responses to the financial crisis. Table 2 lays out the basics of the four bills under study here.⁶

Gramm-Leach-Bliley Act

The first bill and hearings in this study is potentially the most controversial. The Gramm-Leach-Bliley Act repealed key parts of the Glass-Steagall Act of 1933, removing barriers among banks, securities companies and insurance companies that prohibited any institution from behaving as any combination of an investment bank, a commercial bank, and an insurance company. The Gramm-Leach-Bliley Act enabled commercial banks, investment banks, securities firms, and insurance companies to consolidate and engage in increasingly riskier behaviors. It also removed authority, held by the Security and Exchange Commission, to regulate the investing side of primarily lending banks.

Three things make this act significant: it revoked separations and information sharing provisions key to Glass-Steagall, rendering much of the protections from those reforms moot (Posner 2008); it enabled excessively risky investing strategies by commercial banks who, up until now, were prohibited from engaging in the capital raising enterprises of an investment bank, but could now risk private holdings on investment; and it did all this with bipartisan blessing (it was passed by a Republican House and Senate, while being shepherded by President Clinton). This bipartisan framing actually covers something that is much more partisan underneath. In its initial version in the House, it passed on July 1, 1999, by a bipartisan vote of 343-86 (Republicans 205-16; Democrats 138-69; Independent 0-1), but the Senate version (which later became the law) passed on much more partisan lines, 54-44 (53

⁶ This chapter is part of a larger project that, when its complete, will cover eight pieces of key financial legislation, and would introduce four additional bills—the American Recovery and Reinvestment Act and The Housing and Economic Recovery Act to provide two additional post-crisis bills, and the Commodity Futures Modernization Act of 2000 and the Fair and Accurate Credit Transactions Act of 2003 for pre-crisis bills that were discussed as part of the problems leading to the crash.

Republicans and 1 Democrat in favor; 44 Democrats opposed).⁷

The basics behind GLBA, as it is more commonly referred to, had been pursued in the past by Gramm and Bliley. Gramm specifically proposed legislation for rolling back or repealing Glass-Steagall for the first time in the late 1980s—and had the backing of the major banks, the Chairman of the Federal Reserve, Alan Greenspan, and President Bill Clinton, who argued it would “enhance the stability of our financial services system” when signing it into law (Clinton 1999; Crawford 2011). In regards to the bill, two sets of hearings were held, one by the Senate and one by the House, an unusually small number of hearings for such an admittedly important bill. This lack of hearings will play an important role in understanding exactly what happened during the drafting of the bill, but suffice it to say, many important players in shaping the financial environment testified and were active in advocating for their vision of what this act would do.

Sarbanes-Oxley

The next major bill under consideration is also a very critical piece of legislation, Sarbanes-Oxley, passed in response to the major corporate and accounting scandals of the early 2000s, including Enron and WorldCom. The major reforms of Sarbanes-Oxley were designed to curb irregularities in regulation of accounting practices that had allowed major fraud to go undetected for years. Sarbanes-Oxley set new or expanded requirements for all U.S. public company boards, management, and public accounting firms. There are also a number of provisions of the act that also apply to privately held companies, for example the willful destruction of evidence to impede a Federal investigation.

Sarbanes-Oxley significantly changed regulations regarding the reporting of earnings for corporations in the United States, with a particular focus on increasing the public responsibilities of corporate board of directors, in addition to adding criminal penalties for certain

⁷ In something that has become all too prophetic, Congressman John Dingell (D-Mich) argued that Gramm-Leach-Bliley, which was advocated for extensively by representatives of the banking community, would allow banks to become (in his parlance) “too big to fail” by design, and that we risked severe moral hazard problems if the banks expected to be bailed out by the federal government.

malfeasance and requiring the Securities and Exchange Commission to create a new regulatory infrastructure to define, detect, and punish improper reporting of earnings. Sarbanes-Oxley is a truly bipartisan piece of legislation—named after Democratic Senator Paul Sarbanes (D-MD) and Republican Congressman Michael Oxley (R-OH)—and passing the House by a vote of 423-3 and the Senate with a vote of 99-0.

Sarbanes-Oxley created a new regulatory body, the Public Company Accounting Oversight Board (PCAOB), to oversee the practices of accountants at all firms, registering auditors, defining the specific processes and procedures for compliance audits, inspecting and policing conduct and quality control, and enforcing compliance (Piotroski and Srinivasan 2008). While much of the act has been largely accepted, the major regulatory oversight taken by this new regulatory body, and its effect on flow of capital into the US following Sarbanes-Oxley, has been linked to a stronger undercapitalization that exacerbated the financial crisis and potentially lead to more corporate and banking failure than would have otherwise happened (Beyer et al. 2010). Others have criticized the law not for what it does, but for what it *does not do*, namely apply the same strict accounting standards and independence requirements for auditing of corporate funds, to banking and credit rating agencies (Posner 2008). Sarbanes-Oxley had eight hearings over a two week period immediately following the collapse of Enron and these hearings were generally extensive and thorough.

TARP

Now we are moving past bills that lead to (either directly or indirectly) the financial crisis, to the first set of bills designed as a response to the rapidly deteriorating economic conditions. The “Emergency Economic Stabilization Act of 2008”, or TARP as it is more commonly referred to, was written to stave off potential economic collapse by authorizing the Secretary of the Treasury to spend up to \$700 billion to purchase distressed assets (especially the mortgage-backed securities at the heart of the crisis) and in return supply cash directly to the banks.

This bill was selected for two distinct reasons. First, it was passed in a partisan alignment that was almost directly contrary to GLBA (Democratic House, Democratic Senate and Republican President versus Republican House, Republican Senate, and Democratic President) and would provide an interesting contrast along partisan lines. Second, it was the first major bill passed as a response to the financial crisis and in turn gave me an opportunity to study how idea sharing changes during a time of explicit crisis. While there were many hearings over TARP (it was giving out a lot of money, and many banks used the opportunity to ask for money) it almost serves as a control for policy complexity, as the basics of what TARP was doing was relatively straightforward (buying bad assets and distributing money to banks to keep them from failing, essentially a major bail out of the banks). TARP had 16 hearings, numerous speakers, and yet the topic under discussion is not nearly as complicated as the other three bills. Within the “meta-topic” of financial legislation, this proved to be as far a contrast to the larger more complex bills as I could find that still had many prominent public hearings.

Dodd-Frank

Because Dodd-Frank is discussed extensively in Chapter 3 (under the Data section), I will not repeat myself too much here. Suffice it to say, Dodd-Frank was designed to be the most ambitious reform of our major financial laws and regulations since the Great Depression and was written with the above mentioned bills explicitly in mind. Dodd-Frank passed more or less along party lines, with a 223-202 vote in the House and a 59-39 vote in the Senate. What makes Dodd-Frank different than the other bills included here is that the time between the first draft of the bill—December 2009—to the final version—July 2010—was large enough that there were numerous major public hearings. Because this bill was a response to the crisis, and because it was exceedingly ambitious in its purview, the hearings provide a unique opportunity to study the role that outside testifiers have on the writing of legislation. During the seven month period the bill was being considered, there were 63 hearings on elements of

Dodd-Frank, and, as such, every part was explored in extreme depth. The final result is a bill that fuses together the expertise of literally hundreds of speakers, all the while remaining a highly partisan and contentious bill. The partisan aspect to idea borrowing is immense with this bill.

Table 4.1: Major Financial Legislation for chapter

	Section Totals	House	Senate	President	Year Passed
Gramm Leech Bliley	413	R	R	D	1999
Sarbanes Oxley	251	R	D	R	2002
TARP	321	D	D	R	2008
Dodd Frank	2210	D	D	D	2010

Prepping the Data

For any topic modeling approach, we use the bag-of-words approach, meaning that the most efficient way to analyze what a document is saying is to look at the specific term frequencies and tendencies between documents independent of immediate context. Once documents are re-sorted into document-term-matrices, we strip out punctuation, numbers, and “stop words” (words that confer no additional information and are merely linguistic connectors or pronouns). After this, the words are all “stemmed,” taking the word to its most basic, shared root, essentially stripping part-of-speech suffixes.

For this iteration of the project, I fit a single structural topic model on the testimony accompanying the bills and the various drafts of the bills themselves. For the testimony, I treated each uninterrupted speech as a distinct document. All in all, I had 25,000 unique documents. For the bill data, I split the text up into each title for each version of each bill, which gave me a total of 3,100 unique documents. The structural topic model fit on this included a smoothing parameter for time (fit to the prevalence covariate) so the controls of the model would be “dynamic” and, therefore, be flexible enough to change. I also fit a “content covariate” to the model, which was a binary indicator of whether or not the source of the text was from testimony or the bill itself; content covariates are structured

such that they do not overly penalize documents that have covariate specific words that might confuse topic generation (Roberts et al. 2015). Since the writing of bills and the parlance of modern hearings bring with them their own unique vocabulary (a vocabulary that would be unreflected in the substance of the bill section or hearing), this approach ameliorates difference between the two, creating topics that are common to both.

Because of the complexity of the data, and the different document sizes, I fit a model with 74 topics. I used a variety of different k 's for both models before settling on the ones chosen for this chapter: evaluated using the standard model based characteristics described in Stewart, Roberts, and Tingly (2014). Sample fit statistics—including semantic coherence, exclusivity, and held-out-likelihood—are discussed in an appendix and are compared to various other potential settings for k there. I believe that the k 's set for the model is the most coherent models that also fit the various evaluative criteria laid out there.

After fitting the topic model, I run the testimonies through WCopyfind to compare with the bill sections to the relevant bill. From that I count both how many bill sections a given testimony is connected to, and how many phrases they share.

Results

To begin, I limited the analysis to a handful of topics for each model, so as to show the various ways the information here can be explored. After discussing a few interesting topics for both the bills and the testimony, I present the results of modeling the text-reuse algorithm's output. I combine this with other information collected about the speakers to determine what predicts highly "influential" testimonies, and what that reveals about the drafting of these bills.

Table 4.2 presents the results of the topic model described above, with the topic number (generated by the algorithm), the top five words for each topic, and my hand labeling of the topic (not only from reading the top words, but from looking at highly representative documents as well). I excluded the topics that were generally non-policy centric or were a

combination of too many different topics; those are listed in the Appendix. What's left are the policy centric topics which represent the myriad of subjects discussed in the hearings and bill language. Several topics are clearly centered around certain bills and not others. For example, Topic 2, which I've labeled "Accounting Oversight" has, as its second most frequent term, PCAOB, the main regulatory body established by Sarbanes-Oxley.

One can readily see that there are far too many policy centric topics to go through them one at a time and explore how they impact text reuse directly.⁸ What I've done instead is identify four topics from this list that fit the criteria laid out in the theory section for which types of topics should matter—topics that are inherently complex and that deal with implementation—and they are noted in Table 4.2 with a star. These topics, which cover regulating global markets, relation between the Federal Reserve and banks, applications of legal problems, and implementation of inspections and oversight provisions, are diverse enough that they should not overlap in their coverage, nor are they inherently specific to any one bill, but should cover at least a few. The expectation, then, is that each topic will have a strongly positive effect on the instances of textual reuse in legislation.

This approach is different that what was seen in Chapter 3 for two primary reasons. First of all, these bills as was already discussed, both cover very similar subjects broadly but differ largely in the scope of their potential actions and the partisan conditions under which they were passed. So we have something of a varying conditions format to evaluate the theory in. Secondly, by bridging multiple bills, we get around the possibility that the finding in Chapter 3 was unique to Dodd-Frank. If other bills have effects that are similar or greater than Dodd-Frank, then it proves that that bill is not uniquely driving these results.

⁸ A more atheoretical way to go through this would be to use a penalized regression on these variables and see which ones matter and which ones do not, which I've included in an appendix to this chapter. While it won't give us something useful in testing the theories laid out here, it could shed some light on useful ideas for extensions

Table 4.2: Structural Topic Model on Hearings and Bill Sections Policy Topics

Topics	Top Words	Policy Topic
Topic 2	compos, pcaob, staff, collabor, skill	Accounting Oversight
Topic 3	central, global, ampl, disciplin, market	Regulating Global Market*
Topic 4	copay, psychiatr, diagnost, pocket, medic	Medical Insurance
Topic 5	broader, coordin, mission, overview, supervisor	Corporate Organization
Topic 6	nonprim, refinanc, renter, balloon, upfront	Refinancing
Topic 7	subsidiari, depositori, nonmemb, depositor, unsaf	Banking Deposits
Topic 8	decept, unfair, abus, cuomo, consum	Deceptive Practices
Topic 9	gas, cusip, barrel, compress, thereon	Energy/Gas
Topic 10	transcript, complaint, stenograph, forfeit, apportion	Accounting
Topic 11	woman, afflict, actuari, contamin, domenici	Insurance
Topic 12	acquisit, affili, merger, divest, acquir	Mergers and Acquisitions
Topic 15	nasaa, robberi, bookkeep, fidel, advisor	Accounting
Topic 16	depth, complex, style, wide, techniqu	Accounting
Topic 17	overturn, unlicens, annuiti, interchang, reappli	Bank Regulations
Topic 18	atm, recipi, discrep, contractu, disclosur	Bank Regulations
Topic 20	textil, cameron, cell, propuls, counti	Vehicles
Topic 22	thrift, comptrol, abolish, currenc, vest	Currency
Topic 24	commingl, destabil, fund, redempt, estim	Assets
Topic 26	mac, fanni, freddi, fhfa, payout	Fannie and Freddie
Topic 27	interconnected, leverag, pervers, risk, absorb	Risk
Topic 28	tidal, export, rancher, recharg, malpractic	Insurance
Topic 29	reserv, inflationari, mae, feder, doctrin	Federal Reserve *
Topic 30	batteri, daughter, fico, bridg, chain	Insurance
Topic 31	regulatori, deleverag, uniform, streamlin, arbitrag	Regulatory Procedure
Topic 32	wrongdo, turnov, deviat, independ, uneth	Fraud/Bad Practices
Topic 34	overpay, offend, street, wall, portray	Optics/Payments
Topic 35	demograph, girl, outreach, racial, geographi	Demography
Topic 36	council, concentr, diffus, solvent, nonbank	Banking
Topic 37	pecuniari, alleg, suprem, contempt, circuit	Legal/Law*
Topic 39	gallon, truck, motor, hybrid, collar	Vehicles
Topic 40	civic, communiti, neighborhood, stigmat, cra	Communities
Topic 41	monetari, cap, guilti, workout, polit	Payments/Extortions
Topic 42	oil, donat, broker, lesse, deferr	Energy
Topic 43	deputi, vision, appoint, employe, inspector	Inspection and Oversight*
Topic 44	hvcc, reset, collud, subtract, assigne	Home Valuation
Topic 45	product, harbor, ban, pit, unten	Product Regulations
Topic 48	acorn, biolog, english, counselor, softwar	Misc.
Topic 49	uneconom, hedger, counterparti, commod, bid	Hedge Funds
Topic 50	unitari, tradabl, hold, mutual, undercapit	Capitalization
Topic 51	frivol, scienter, reinstat, aider, civil	Lawsuits
Topic 53	countercycl, tranch, auction, asset, collater	Assets and Transactions
Topic 55	tabul, compil, readabl, obey, store	Data
Topic 56	reloc, bankruptci, autopsi, whoever, implicit	Bankruptcy
Topic 57	planner, exposur, ventur, conting, steagal	Venture Capital
Topic 58	supervisor, examin, forg, deleg, fullest	Financial Planning
Topic 59	research, diverg, circumv, defect, client	Research
Topic 65	nomine, reincorpor, proxi, plural, cast	Accounting
Topic 66	mccarran, reinsur, naic, barnett, preempt	Insurance
Topic 67	swap, pari, dealer, circul, major	Swaps
Topic 68	price, risen, cost, valuat, overhead	Costs
Topic 69	abolit, proportion, life, basic, tenur	Insurance
Topic 73	methodolog, grade, plead, obligor, rate	Credit Ratings

Test-Reuse Results

I ran WCopyfind on testimonies and the bill sections, with manual cleaning of the matches identified by the algorithm, to eliminate false positives. Below are the results of counting the number of bill sections to which a given testimony is connected.

Consider, first, the relative rates by the various bills. Table 4.3 lays this out in nice detail. What is abundantly clear from this initial analysis is that there are marked differences between how frequently hearings over a given bill utilize language that later ends up in the bill itself. Table 4.3 shows not only how many hearings there were for each bill but also how many speeches there were, laying out a baseline for what reuse rates we should expect. When looking at the last two columns, which are two different ways of assessing how much textual reuse is occurring between the hearings and these specific bills, a rather different set of outcomes emerges. While Gramm-Leach-Bliley, Sarbanes-Oxley, and Dodd-Frank all have relatively high levels of reuse for their speeches, TARP does not. TARP, while having almost ten times the amount of speeches during their hearings as Gramm-Leach-Bliley, only has 49 speeches where the algorithm detected text reuse, whereas Gramm-Leach-Bliley has 40. This finding fits in the general expectation that TARP would have fewer instances of idea borrowing than the other legislation because the task it was undertaking had little to do with implementation and was, at its heart, a rather straightforward task (i.e., giving the Secretary of the Treasury the ability to buy up bad assets for banks that were in danger of going bankrupt).

Table 4.3: Major Financial Legislation and Reuse Rates

	Total Hearings	Speeches	Speeches w/Reuse	Total Reused Ideas
Gramm Leech Bliley	2	244	40	11718
Sarbanes Oxley	8	838	70	17838
TARP	12	2076	49	2183
Dodd Frank	63	7353	500	158113

To best understand what predicts a testimony sharing content with a bill section, I

decided to model different forms of text reuse on the topics of interest. I ran three distinct models, all using a different conceptualization of text reuse from the data: a Logit model on whether or not a given piece of testimony has any connections at all, a negative binomial model on how many bill sections a given testimony is connected to, and another negative binomial model on how many overall connections a given piece of testimony has. I selected the four potentially significant topics outlined above (Topics 3, 29, 37, and 43) and added in the the source of the hearings (the four bills) as either an additional covariate (in the first three models) or as a random intercept in the mixed effect models (the last three models). Table 4.4 lays out each of these models.

The results of these models are clear in several regards. First of all, regarding the non-hierarchical models overall, there are some clear differences between the four bill sources as a basis for the regressions. The model uses Dodd-Frank as the reference point, so it is not presented in the table. Consistently, TARP is negatively associated with all three different conceptualizations of text reuse, Gramm-Leach-Bliley is positive (but sometimes not statistically significant), and Sarbanes-Oxley is inconsistent.

Beyond this, however, we see a very strong relationship between the chosen topics. As predicted, all four have a very strongly positive correlation to higher levels of text reuse. Consistently, the strongest predictor of reuse in all forms is the Oversight topic, which makes sense if we think about idea borrowing as the clearest manifestation of sharing in expertise. Since oversight is driven by discussion of implementation and how Congress (or the agencies they set up) are to perform this function, those experts speaking about this have the strongest priors as to what future oversight should look like. Shaping exactly how this comes together would be an important role to be explicated during these hearings, so it is unsurprising that the greatest degree of idea borrowing occurs there.

Figures 4.1 and 4.2 explore this more in-depth by showing the predicted text reuse counts over the topic of choice (all while holding the other topics at their mean, and faceting over the four different bills). Figure 4.1 shows the dramatic increase in estimated text reuse sections

Table 4.4: Modeling Text Reuse of Financial Legislation with Select Topics

	<i>Dependent variable:</i>					
	reuse <i>Logit</i> (1)	count <i>glm: NB</i> <i>link = log</i> (2)	total <i>glm: NB</i> <i>link = log</i> (3)	reuse <i>Logit</i> <i>mixed-effects</i> (4)	count <i>NB</i> <i>mixed-effects</i> (5)	total <i>NB</i> <i>mixed-effects</i> (6)
Federal Reserve	10.052*** (1.031)	17.620*** (2.544)	12.823*** (2.806)	10.118*** (1.030)	17.625*** (0.548)	20.424*** (0.514)
Regulations	14.099*** (1.189)	17.997*** (3.052)	11.980*** (3.262)	14.078*** (1.189)	17.999*** (0.548)	18.709*** (0.479)
Oversight	17.621*** (2.486)	51.163*** (6.045)	135.817*** (6.700)	17.554*** (2.486)	51.133*** (1.635)	66.187*** (1.633)
Legal/Law	15.220*** (2.689)	27.544*** (6.429)	56.172*** (6.828)	15.460*** (2.704)	27.575*** (1.646)	55.916*** (1.731)
GLBA	0.887*** (0.189)	0.728 (0.474)	0.464 (0.507)			
SOX	0.508*** (0.142)	0.120 (0.330)	-1.075*** (0.309)			
TARP	-0.824*** (0.155)	-2.650*** (0.521)	-2.998*** (0.278)			
Obs	10,511	10,511	10,511	10,511	10,511	10,511
AIC	4,593	26,353	68,561	4,607	26,382	66,178
BIC				4,650	26,433	66,229

Note:

*p<0.1; **p<0.05; ***p<0.01

by increasing only the probability for the oversight topic. Increasing it by a percentage, for some bills, moves it up by thousands of counts, a very strong connection. In contrast, Figure 4.2 shows the predicted text reuse counts over the regulatory topic, which is generally more common, and results in only an additional few hundred reused sections as we increase percent by percent. The difference, however, between the four bills is striking, with GLBA and Dodd-Frank always having the most reuse and TARP having far and away the least.

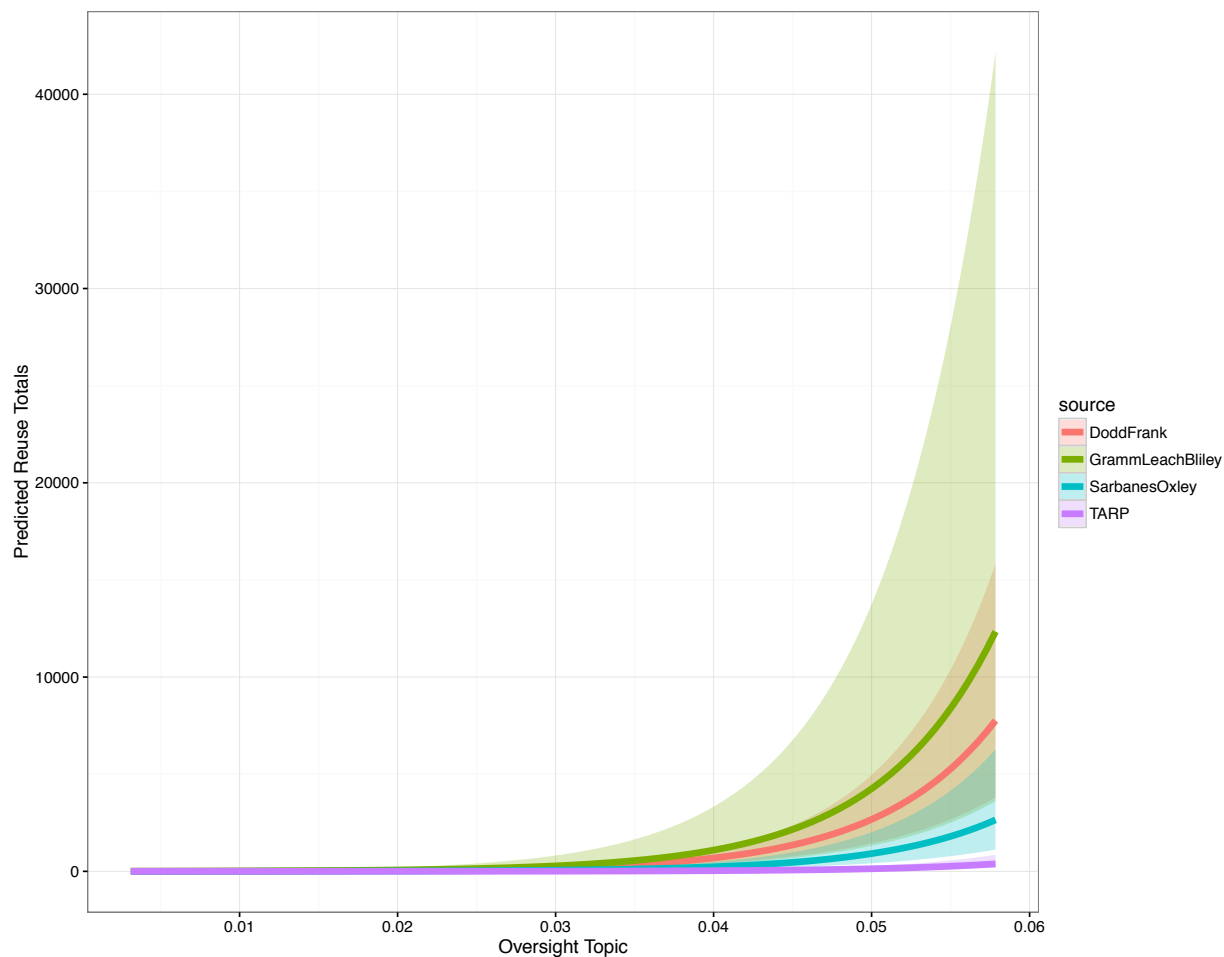


FIGURE 4.1: Predicted Counts by Oversight Topic over Bill

Conclusion and Next Steps

Chapter 4 reinforces the importance that the text of bills—namely the ideas that influence and later become the written law—should have when we discuss Congress. In this chapter, I have demonstrated the ways in which Congress engages in idea borrowing, and how that impacts the writing of law. Looking at the role of major financial legislation, we are able to see how certain topics predict differential congressional behavior, suggesting an new avenue for a research agenda getting into the bill writing process.

This chapter is very clearly part of a greater project on how the writing of financial legislation came about. Future iterations of this chapter will include individual level covariates

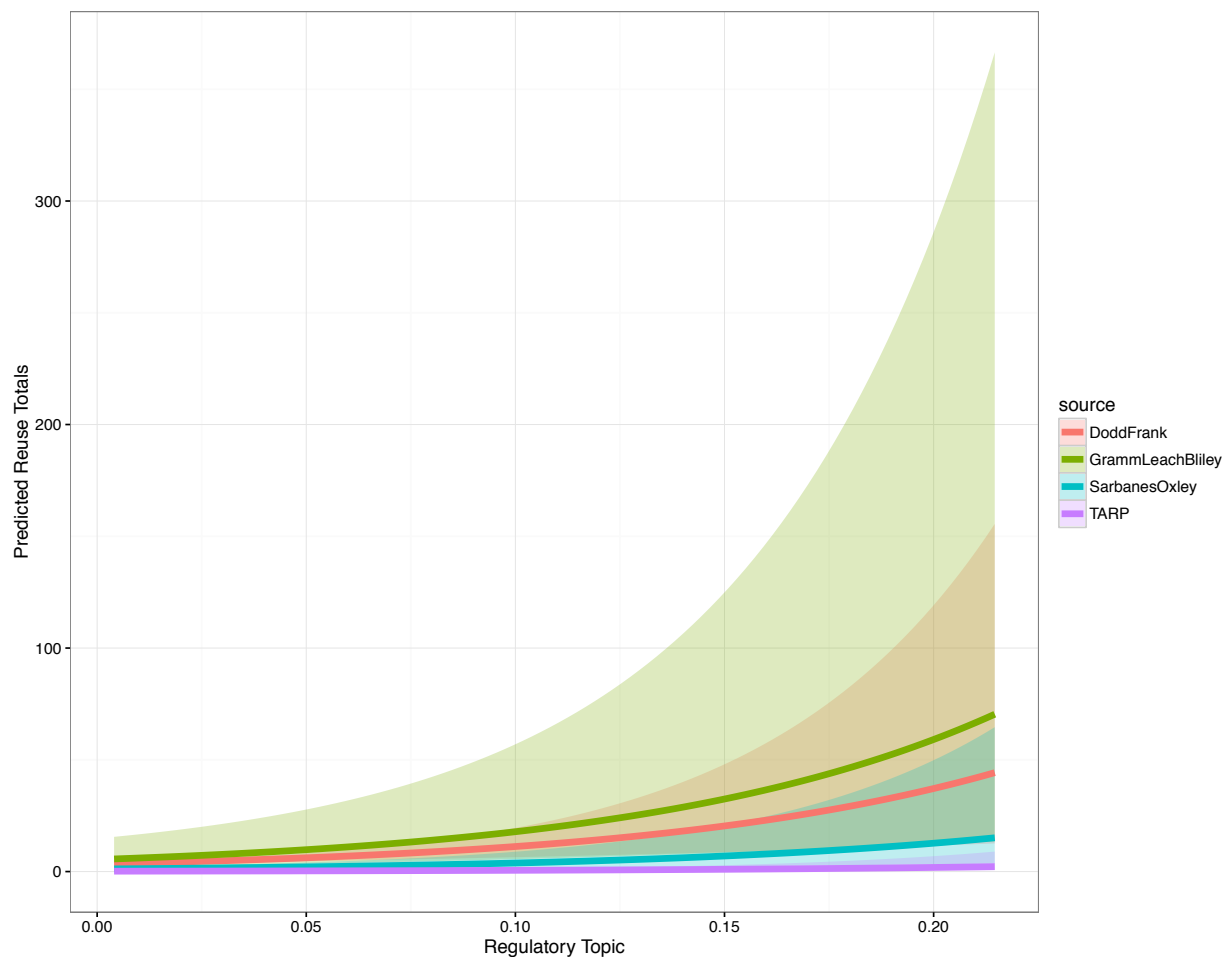


FIGURE 4.2: Predicted Counts by Regulatory Topic over Bill

for the speakers that will get into exactly who does Congress trust to testify and who do they listen to. This is probably the biggest weakness of the chapter, because the coding of the individual speaker level covariates will add in a layer to the analysis that should be able to strengthen the external validity of the model. This will also address the question of whom Congress trusts in particular, and it will create the opportunity to explore who exactly testifies at these hearings as well, let alone who is listened to.

Future work will also address in a more robust way additional legislation, some less significant than the ones discussed above, so more reasonable inference about the conditions under which borrowing occurs can be made. This chapter broadly introduces a theory and data set that now require a number of smaller tests that must be run in order to get a better

sense of exactly what is the relevant treatment predicting idea borrowing through text reuse. To get at these ideas (issues such as the effect of different partisan arrangements), we will need to perform this operation on bills that are nearly identical. Indeed, this process of selecting just financial legislation gets at that indirectly by focusing the study to just similar bill subjects. Future work could prioritize the development of a more formal counterfactual to better study this effect.

Regardless, this chapter was able to get at the fundamental question about where ideas in legislation come from. By utilizing these major pieces of financial legislation, this chapter concludes that it is only when the subject of a hearing has to do with complex and implementation centric issues that Congress will directly borrow ideas from testimonies. Understanding this gets us closer to understanding exactly where ideas in Congress come from, and what impact they have on the law.

5

Conclusion

In the dime stores and bus stations
People talk of situations
Read books, repeat quotations
Draw conclusions on the wall

Bob Dylan

Even after analysis of all this material, it is still clear that the role that ideas play in politics remains unsatisfactory. The findings presented in this project attempts to bridge that gap and to add to the already growing corpus of work utilizing text-as-data to study Congress. These papers present together a comprehensive whole that makes some larger arguments more broadly. Even though the papers are bound less tightly by topic than I would have liked, the overarching argument made herein remains coherent and significant.

The significant conclusions from these papers are as follows: context matters, ideology matters, substance matters, and, above all else, ideas matter. Think tanks see themselves not just as research organs or as vehicles to express a few policy opinions publicly but as parts of a larger ideological whole, strategically positioning themselves such that they can maximize their influence while maintaining both ideological purity and the appearance of objectivity. Congressmen, when listening to testimony during contentious hearings over

important legislation, will “repeat quotations” and engage in idea reuse when writing the bills. What makes it into the final bills, then, is the product of the collection of ideas from the various people called to testify, in addition to whatever MCs start with and add to during the legislative process. To that point, this isn’t to argue that the ideas tracked in this work are exclusive, nor is it entirely comprehensive, since the ideas can come from anywhere, but it is to say that ideas transmitted in hearings do matter, that the context in which words are spoken change how and when MCs listen, and indeed, that those situations lead to the writing of different bills.

The possibilities for future work beyond this is nearly endless. Once one begins to unwrap the ideas behind important legislation, the exceedingly important task is to trace it back and forth between Congress and the general public. Modeling this in a comprehensive way could revolutionize how we view political discourse and how legislation is understood. Additional work extending these techniques into the finer details of the legislative process could unearth rather important truths about how law is structured. Brining the big-data revolution to the study of legislative institutions and procedures is vital to keeping this research at the forefront. Producing a more comprehensive map of legislative ideas could be the bridge needed between important systematic quantitative work on Congressional behavior and intuitive and nuanced qualitative studies.

Appendix A

Extras to Chapter 2

A.1 Appendix: More information about Board of Directors of Think Tanks

The collecting of information about the different Boards of Directors formed the basis for the discussions in this paper about the ideologies of the aforementioned think tanks. This process involved collecting information about each board member from the various websites for the organizations, matching that record to individual records at the FEC tracking websites, particularly from the Center for Responsive Politics’ “Open Secrets” website, which gave a relatively straightforward way of searching through contribution records for individuals as well as corporations. Once I identified the person in CRP’s database of individual contributions, I matched it to Adam Bonica’s database of individual contributors, based on name, sex, residency, and stated occupation¹. Overall, I was able to weakly identify 1892 out of a possible 2610 and strongly identify 1636 total. I only ended up using the strongly

¹ Which was somewhat problematic in and of itself, because individuals were not constrained as to what they would list as their professions. An example: Richard Mellon Scaife, the prominent conservative philanthropist and businessman, who is on the board of the Heritage Foundation, listed on separate contribution forms both “Businessman” and “Chairman” as his professions. While this is obviously the same thing in this case, it made machine matching the cases much more difficult.

Table A.1: Distribution of CFScore for Think Tanks

Min	1st Quartile	Median	Mean	3rd Quartile	Max	N
-1.71	-.0852	0.15	0.036	0.918	2.4	1636

identified cases, but there were no systematic differences in the organization level scores, so it should not be a concern. Here are the summary statistics for the CFScores of the board members included in this dataset.

There were a few individual board members that had scores well outside the range for any of the organizations, or of any Congressmen, but for the most part individual scores mapped nicely to organization scores, again lending credence to the representative nature of these scores.

Overall board size varied tremendously, with several boards simply being four or fewer members, and several having board sizes of over 50. The smaller boards tended to get excluded from final analysis, unless they had nearly complete membership online; several organizations are essentially headed by one prominent person, who were always found in the dataset, so those board summary scores were included since they seem to have a board director as organizational dictator leadership model.

A.2 Appendix: Different Count Regression Methods

The different model specifications and method selection does very little to alter the substantive results discussed before. I used the Bayesian models for the main part of the paper because of the lower sample size, dataset complexity, and the tendency for Bayesian models to more properly deal with discrepancies with over dispersion (which is the reason to use a negative binomial regression in the first place). Since there is minimal substantive differences between the results reported here, and the results interpreted in the paper, it is probably safe to assume that the model chosen is correct.

Table A.2: Alternative Citations Models

	Bayesian Poisson	Negative Binomial	Poisson
(Intercept)	2.58*** (0.46)	2.71*** (0.33)	2.56*** (0.04)
Employees	0.03** (0.01)	0.05** (0.02)	0.03*** (0.00)
ABS CFscore2	1.20** (0.44)	0.85* (0.34)	1.22*** (0.04)
Age in 2005	0.01 [†] (0.00)	0.01 [†] (0.00)	0.01*** (0.00)
Lobbyists	0.04 (0.08)	0.06 (0.09)	0.04*** (0.01)
Expenditures by \$1000	0.00*** (0.00)	0.00* (0.00)	0.00*** (0.00)
N	125	125	125
AIC	10006.27	1281.88	10016.72
BIC	10074.15	1361.07	10081.11
$\log L$	-4979.13	-612.94	-4988.31

Standard errors in parentheses

[†] significant at $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

For the Testimony models, the same results hold up to scrutiny: both CFscore and Employees remain significant, and, although different controls go in and out of mattering to the overall result, the actual substantive interpretations one could make from these models remains unchanged.

Table A.3: Alternative Testimony Models

	Bayesian Poisson	Negative Binomial	Poisson
(Intercept)	4.05*** (0.03)	4.08*** (0.31)	4.05*** (0.03)
ttdata.impute\$Employees	0.05*** (0.00)	0.06*** (0.01)	0.05*** (0.00)
ttdata.impute\$cfscore2	-0.41*** (0.03)	-0.48 (0.32)	-0.41*** (0.03)
ttdata.impute\$age2005	0.00*** (0.00)	-0.00 (0.00)	0.00*** (0.00)
ttdata.impute\$Lobbyists	-0.00 (0.01)	0.04 (0.09)	-0.00 (0.01)
ttdata.impute\$Expend1000	0.00*** (0.00)	0.00*** (0.00)	0.00*** (0.00)
N	125	125	125
AIC	10543.71	1336.50	10552.11
BIC	10611.59	1415.69	10441.23
$\log L$	-5247.85	-640.25	-5242.51

Standard errors in parentheses

† significant at $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

With that in mind, I also extended the citations models to allow them to vary Congress to Congress. Given how the testimony data was collected, I was unable to run the same models on both different dependent variables². For this data, I use a zero-inflated model and a hurdle model, both specified for working with count data with too many zeros.

Substantively, the biggest result from this is how much greater an impact every control variable has when a dummy variable for the Congress the citations occurred during is included; the 107th Congress has an unusually strong (negative) effect on total citations. Much like the previous regression tables, in this table there is very little difference between the Negative Binomial cases and the Poisson cases.

² The Testimonies data were just collected in a single large chunk for each organization, while the citations were done Congress by Congress.

Table A.4: Citations by Year per Think Tank

	<i>Dependent variable:</i>			
	Citations by Year			
	<i>zero-inflated count data</i>		<i>hurdle</i>	
	Negative Binomial	Poisson	Negative Binomial	Poisson
ABS CFscore	0.886*** (0.165)	1.021*** (0.037)	0.917*** (0.196)	1.022*** (0.037)
Age	0.008*** (0.003)	0.006*** (0.0004)	0.008*** (0.003)	0.006*** (0.0004)
Lobbyists	0.133*** (0.049)	0.065*** (0.006)	0.131** (0.056)	0.065*** (0.006)
Employees	0.044*** (0.010)	0.018*** (0.001)	0.038*** (0.011)	0.018*** (0.001)
Expenditures	0.145*** (0.038)	0.219*** (0.008)	0.142*** (0.044)	0.219*** (0.008)
104th Congress	0.039 (0.171)	0.029 (0.029)	0.041 (0.201)	0.029 (0.029)
105th Congress	-0.279 (0.171)	-0.171*** (0.031)	-0.238 (0.204)	-0.171*** (0.031)
106th Congress	-0.239 (0.169)	-0.196*** (0.030)	-0.264 (0.199)	-0.196*** (0.030)
107th Congress	-0.794*** (0.172)	-0.620*** (0.035)	-0.821*** (0.205)	-0.620*** (0.035)
Constant	0.266 (0.348)	-0.085 (0.085)	0.180 (0.416)	-0.087 (0.085)
Observations	625	625	625	625
Log Likelihood	-2,144.027	-6,421.407	-2,135.636	-6,421.196

Note:

*p<0.1; **p<0.05; ***p<0.01

Appendix B

Extras to Chapter 3

Top Topics for Testimony

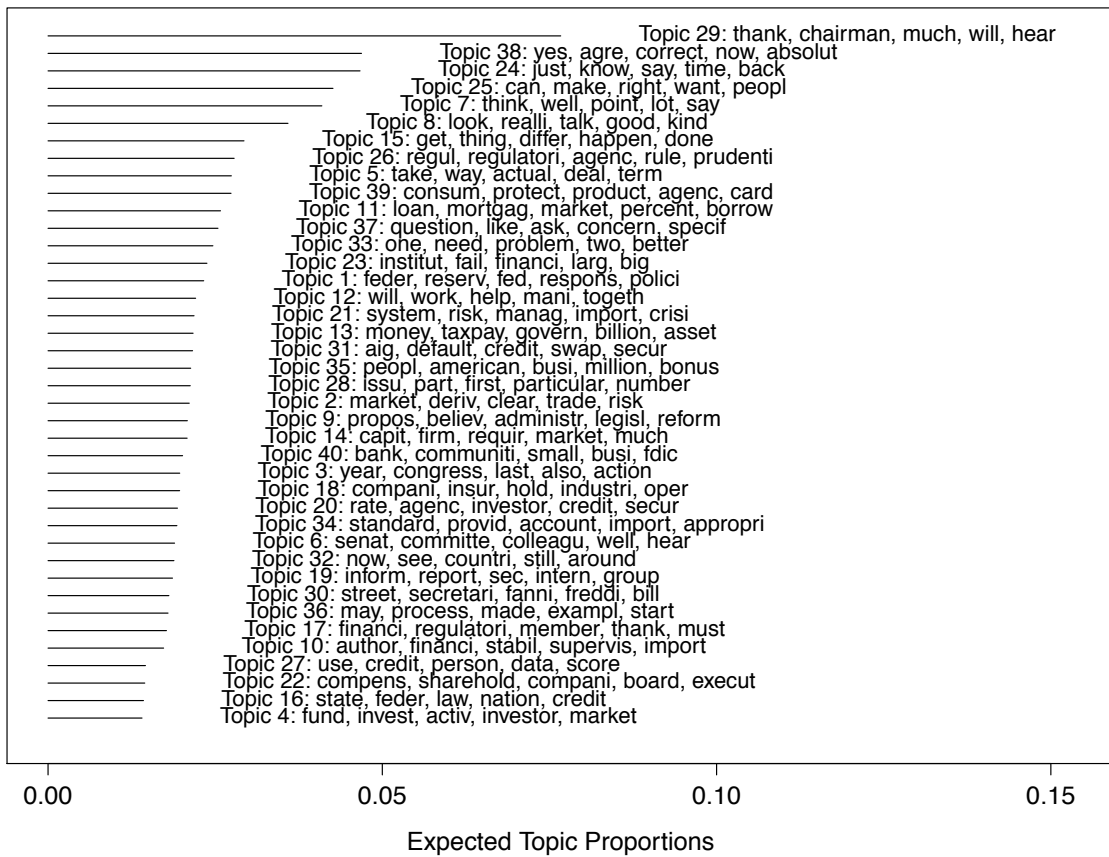


FIGURE B.1: Topic Model for Testimony Only

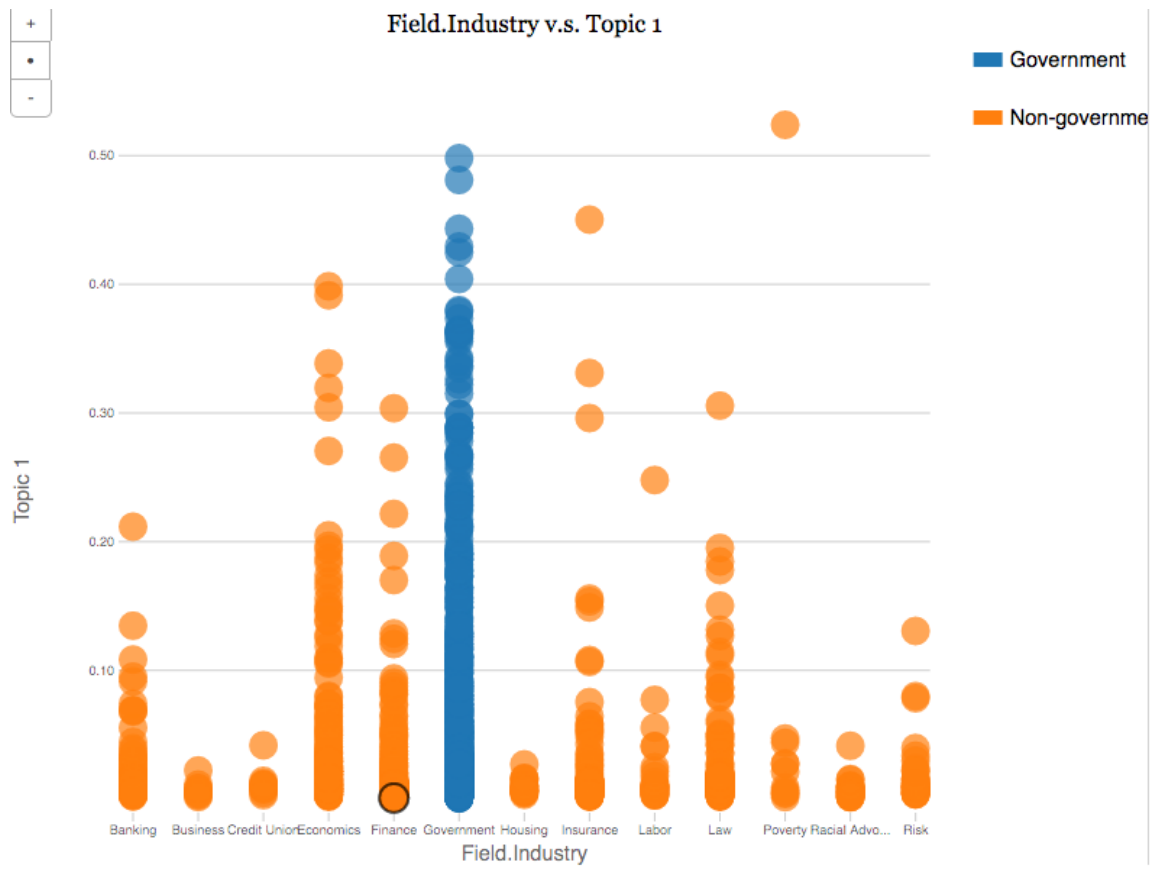


FIGURE B.2: Changes to Federal Reserve Topic

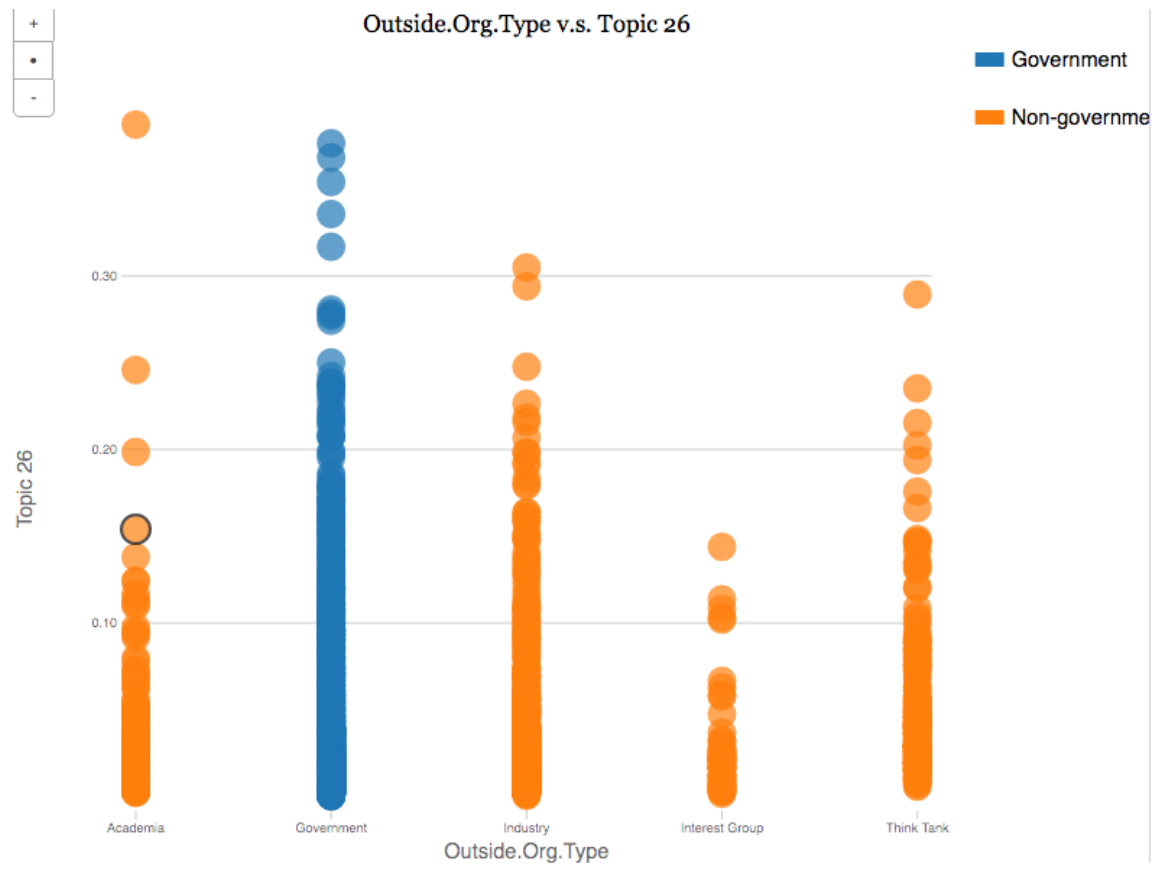


FIGURE B.3: Changes to Regulation and Rules Topic

Top Topics for Text of Bill

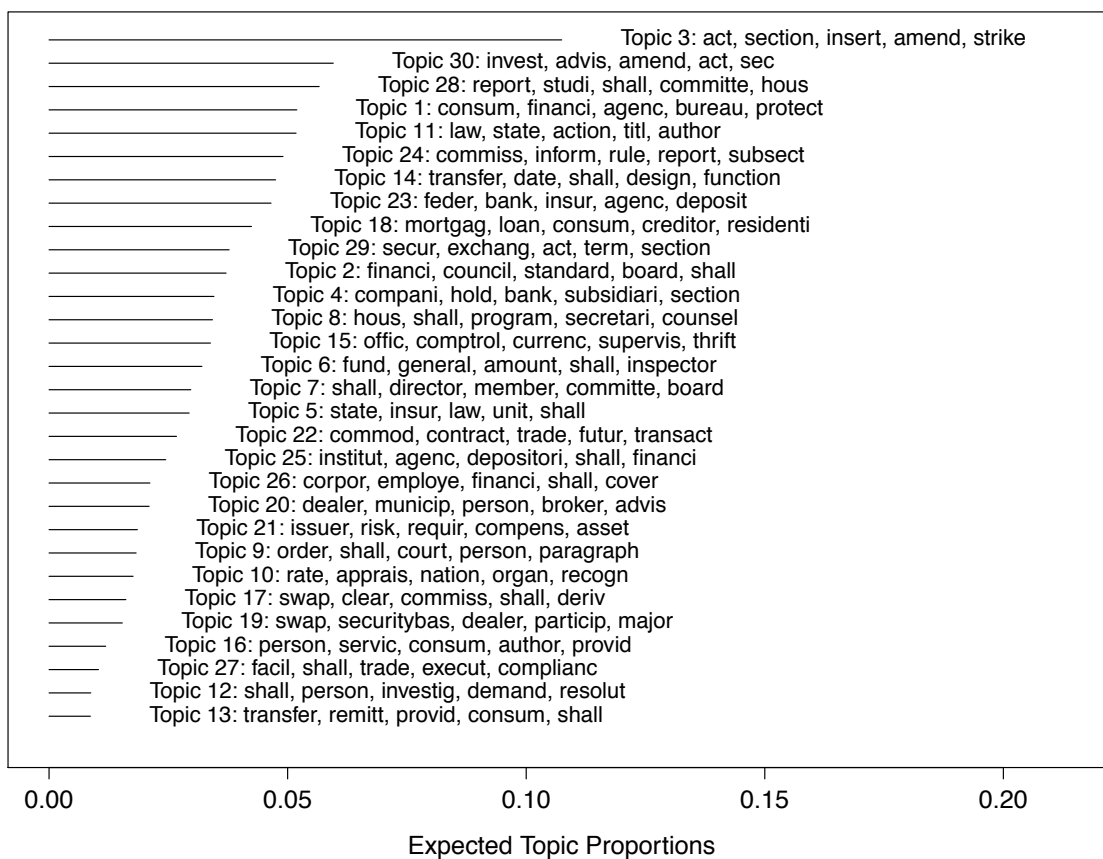


FIGURE B.4: Topic Model for Bill Only

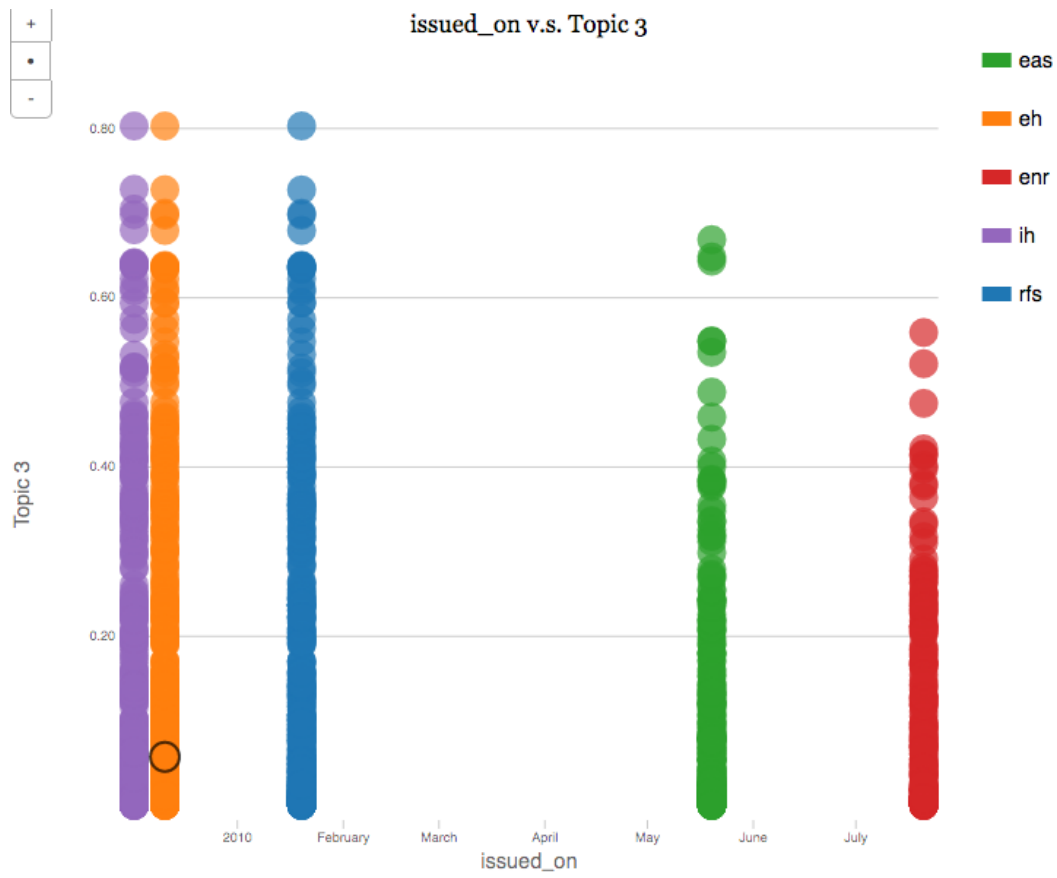


FIGURE B.5: Changes to Amending Topic

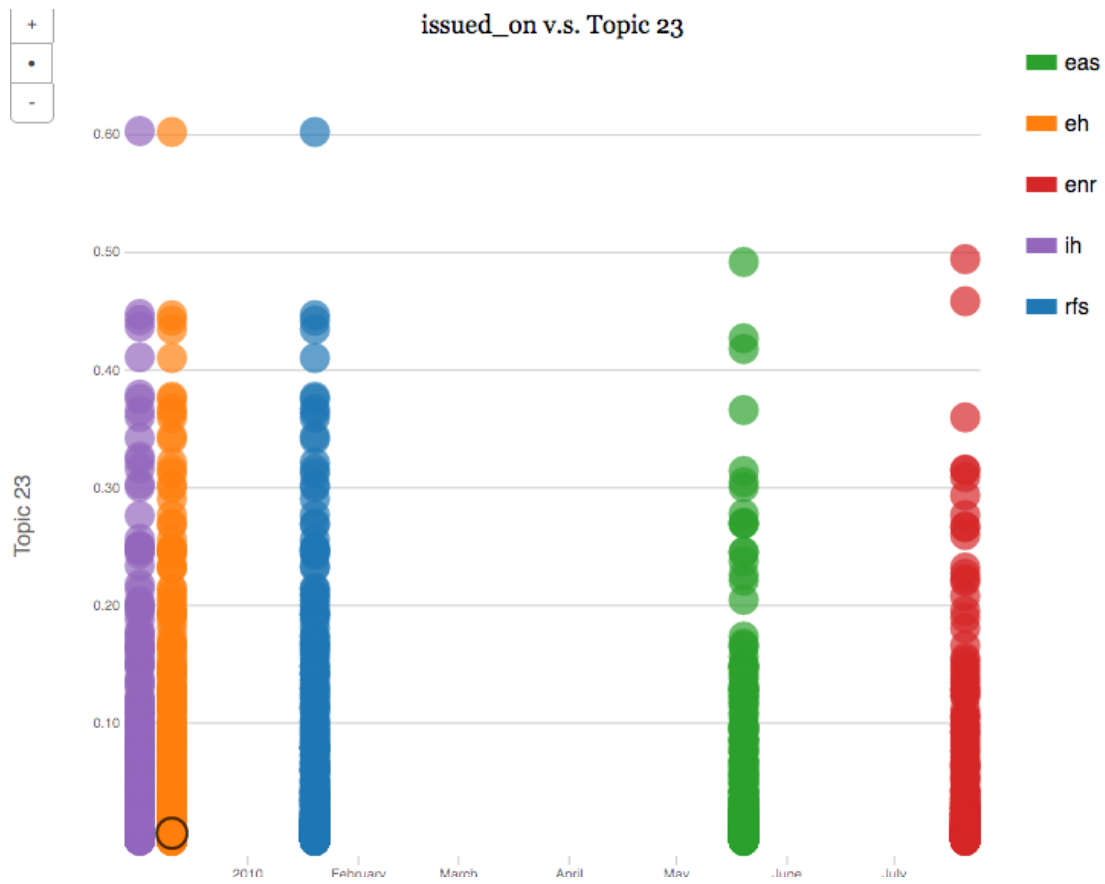


FIGURE B.6: Changes to Federal Banking Topic

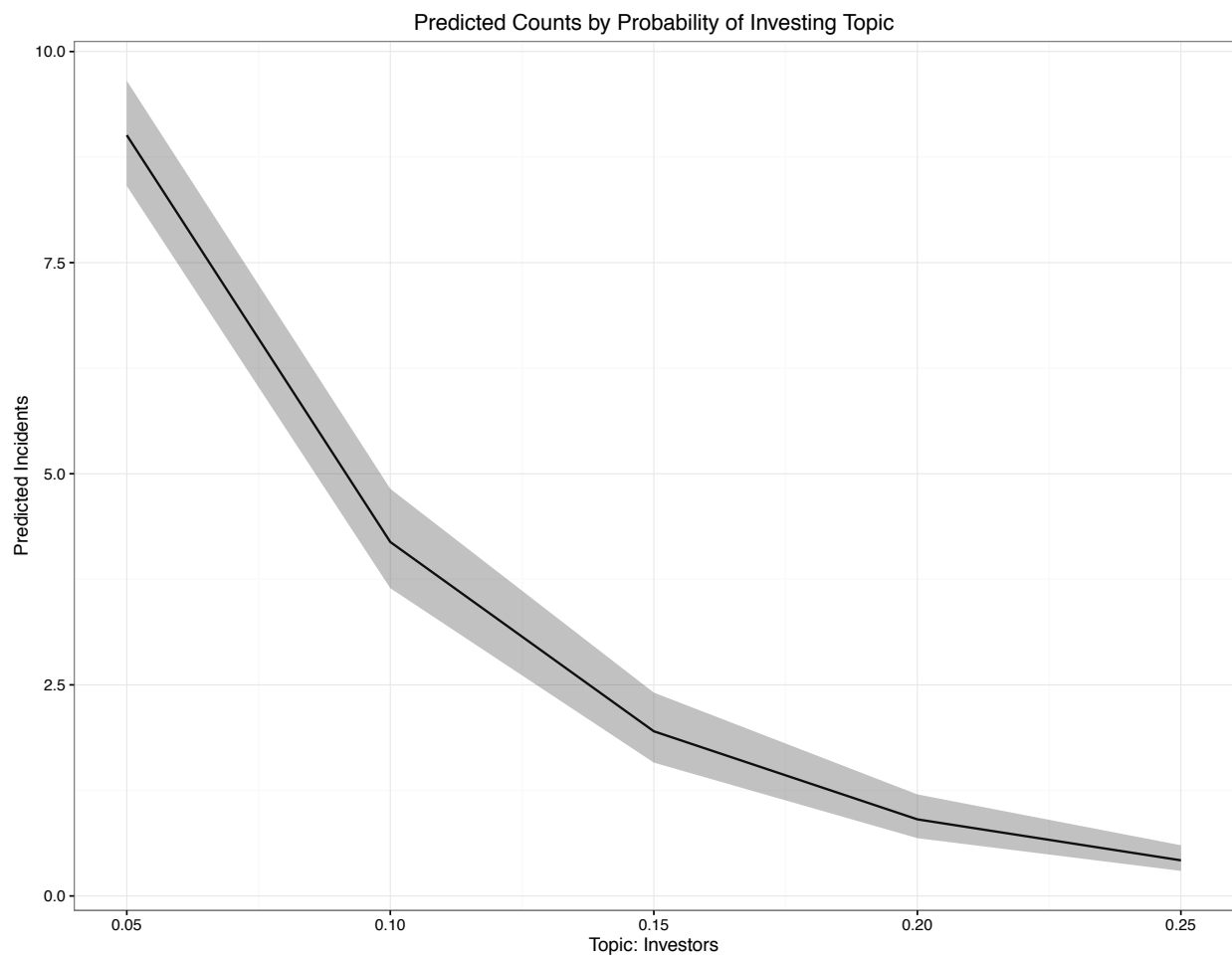


FIGURE B.7: Alternative Specification of Topics: Invest Topic

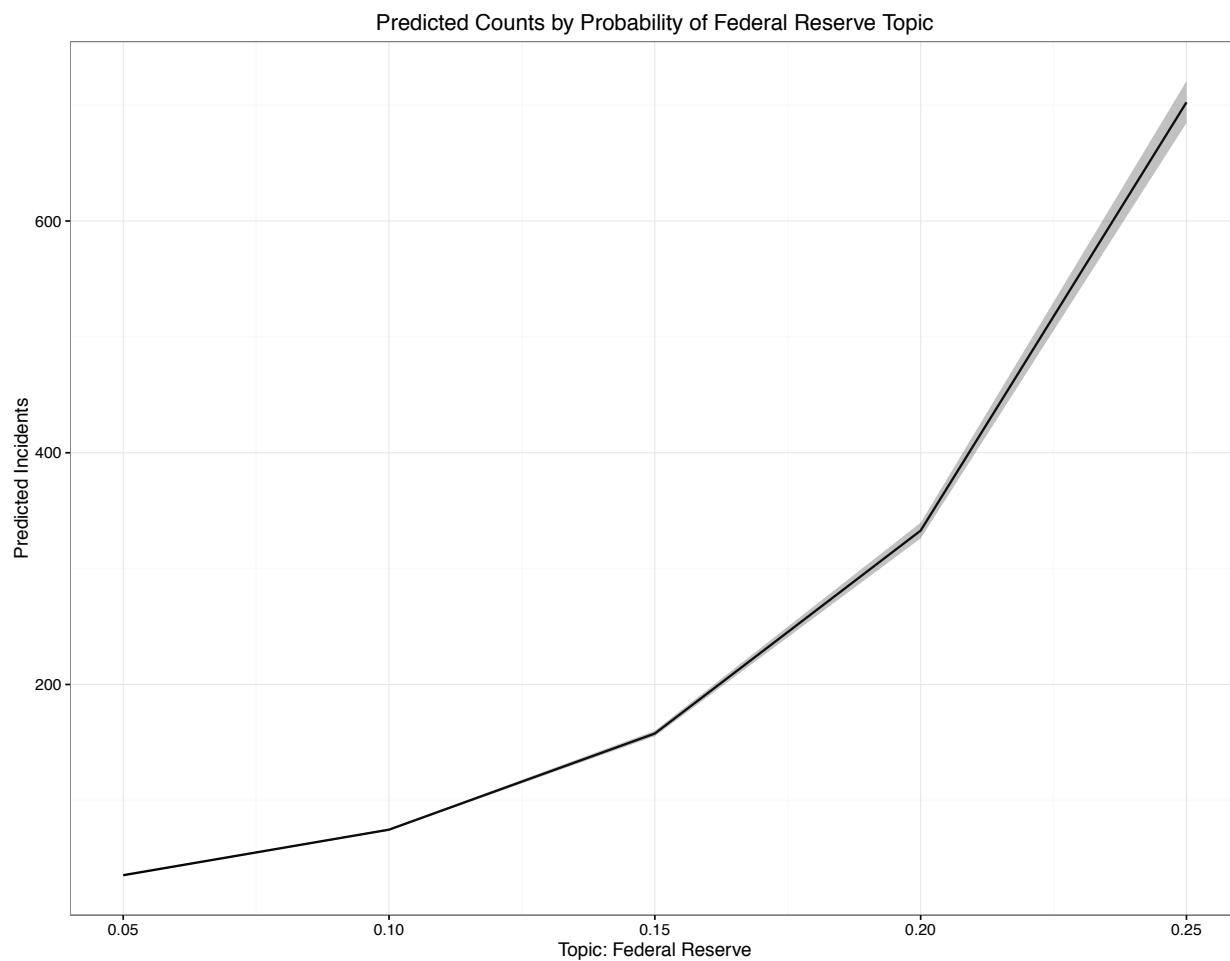


FIGURE B.8: Alternative Specification of Topics: Federal Reserve Topic

Appendix C

Extras to Chapter 4

Table C.1: Structural Topic Model on Hearings and Bill Sections w/74 Topics

Topics	Top Words	Policy Topic
Topic 1	gene, concern, observ, whole, forward	None
Topic 2	compos, pcaob, staff, collabor, skill	Accounting Practices
Topic 3	central, global, ampl, disciplin, market	Regulating Global Market*
Topic 4	copay, psychiatr, diagnost, pocket, medic	Medical Insurance
Topic 5	broader, coordin, mission, overview, supervisor	Corporate Organization
Topic 6	nonprim, refinanc, renter, balloon, upfront	Refinancing
Topic 7	subsidiari, depository, nonmemb, depositor, unsaf	Banking Deposits
Topic 8	decept, unfair, abus, cuomo, consum	Deceptive Practices
Topic 9	gas, cusip, barrel, compress, thereon	Energy/Gas
Topic 10	transcript, complaint, stenograph, forfeit, apportion	Accounting
Topic 11	woman, afflict, actuari, contamin, domenici	Insurance
Topic 12	acquisit, affili, merger, divest, acquir	Mergers and Acquisitions
Topic 13	whatsoev, presum, rout, game, forth	None
Topic 14	interlock, sentenc, surfac, strike, blank	None
Topic 15	nasaa, robberi, bookkeep, fidel, advisor	Accounting
Topic 16	depth, complex, style, wide, techniqu	Accounting
Topic 17	overturn, unlicens, annuiti, interchang, reappli	Bank Regulations
Topic 18	atm, recipi, discrep, contractu, disclosur	Bank Regulations
Topic 19	bliley, gramm, fingerprint, master, canadian	None
Topic 20	textil, cameron, cell, propuls, counti	Vehicles
Topic 21	togeth, probabl, know, insid, afield	None
Topic 22	thrift, comptrol, abolish, currenc, vest	Currency
Topic 23	democrat, treasur, begin, fbi, seri	None
Topic 24	commingl, destabil, fund, redempt, estim	Assets
Topic 25	analyst, restat, acquiesc, embodi, arisen	None
Topic 26	mac, fanni, freddi, fhfa, payout	Fannie and Freddie
Topic 27	interconnected, leverag, pervers, risk, absorb	Risk
Topic 28	tidal, export, rancher, recharg, malpractic	Insurance
Topic 29	reserv, inflationari, mae, feder, doctrin	Federal Reserve*
Topic 30	batteri, daughter, fico, bridg, chain	Insurance
Topic 31	regulatori, deleverag, uniform, streamlin, arbitrag	Regulatory Procedure
Topic 32	wrongdo, turnov, deviat, independ, uneth	Fraud/Bad Practices
Topic 33	cfpa, safeti, sound, preval, arena	None
Topic 34	overpay, offend, street, wall, portray	Optics/Payments
Topic 35	demograph, girl, outreach, racial, geographi	Demography
Topic 36	council, concentr, diffus, solvent, nonbank	Banking
Topic 37	pecuniari, alleg, suprem, contempt, circuit	Legal/Law
Topic 38	cloth, withdrew, recycl, inch, claim	None
Topic 39	gallon, truck, motor, hybrid, collar	Vehicles
Topic 40	civic, communiti, neighborhood, stigmat, cra	Communities
Topic 41	monetari, cap, guilti, workout, polit	Payments/Extortions
Topic 42	oil, donat, broker, lesse, deferr	Energy
Topic 43	deputi, vision, appoint, employe, inspector	Inspection and Oversight*
Topic 44	hvcc, reset, collud, subtract, assigne	Home Valuation
Topic 45	product, harbor, ban, pit, unten	Product Regulations
Topic 46	speci, inner, card, plastic, minus	None
Topic 47	falsifi, audit, disciplinari, ethic, forma	None
Topic 48	acorn, biolog, english, counselor, softwar	Misc.
Topic 49	uneconom, hedger, counterparti, commod, bid	Hedge Funds
Topic 50	unitari, tradabl, hold, mutual, undercapit	Capitalization
Topic 51	frivol, scienter, reinstat, aider, civil	Lawsuits
Topic 52	quotat, denot, unconsolid, nonpubl, deter	None
Topic 53	countercycl, tranch, auction, asset, collater	Assets and Transactions
Topic 54	berni, consent, correct, right, statutorili	None
Topic 55	tabul, compil, readabl, obey, store	Data
Topic 56	reloc, bankruptci, autopsi, whoever, implicit	Bankruptcy
Topic 57	planner, exposur, ventur, conting, steagal	Venture Capital
Topic 58	supervisor, examin, forg, deleg, fullest	Financial Planning
Topic 59	research, diverg, circumv, defect, client	Research
Topic 60	quorum, terminolog, speaker, veto, introduct	None
Topic 61	teaser, african, livelihood, yet, tank	Misc.
Topic 62	answer, interrupt, propound, oath, china	None
Topic 63	hunter, expir, clerk, miller, campus	None
Topic 64	forthwith, thereupon, desist, untru, recur	None
Topic 65	nomine, reincorpor, proxi, plural, cast	Accounting
Topic 66	mccarran, reinsur, naic, barnett, preempt	Insurance
Topic 67	swap, pari, dealer, circl, major	Swaps
Topic 68	price, risen, cost, valuat, overhead	Costs
Topic 69	abolit, proportion, life, basic, tenur	Insurance
Topic 70	studi, analyz, compact, feasibl, coerc	None
Topic 71	secretari, treasuri, sent, chairman, weed	None
Topic 72	succinct, phrase, make, judg, problem	None
Topic 73	methodolog, grade, plead, obligor, rate	Credit Ratings
Topic 74	atmospher, metric, book, foolish, everi	None

Diagnostic Values by Number of Topics

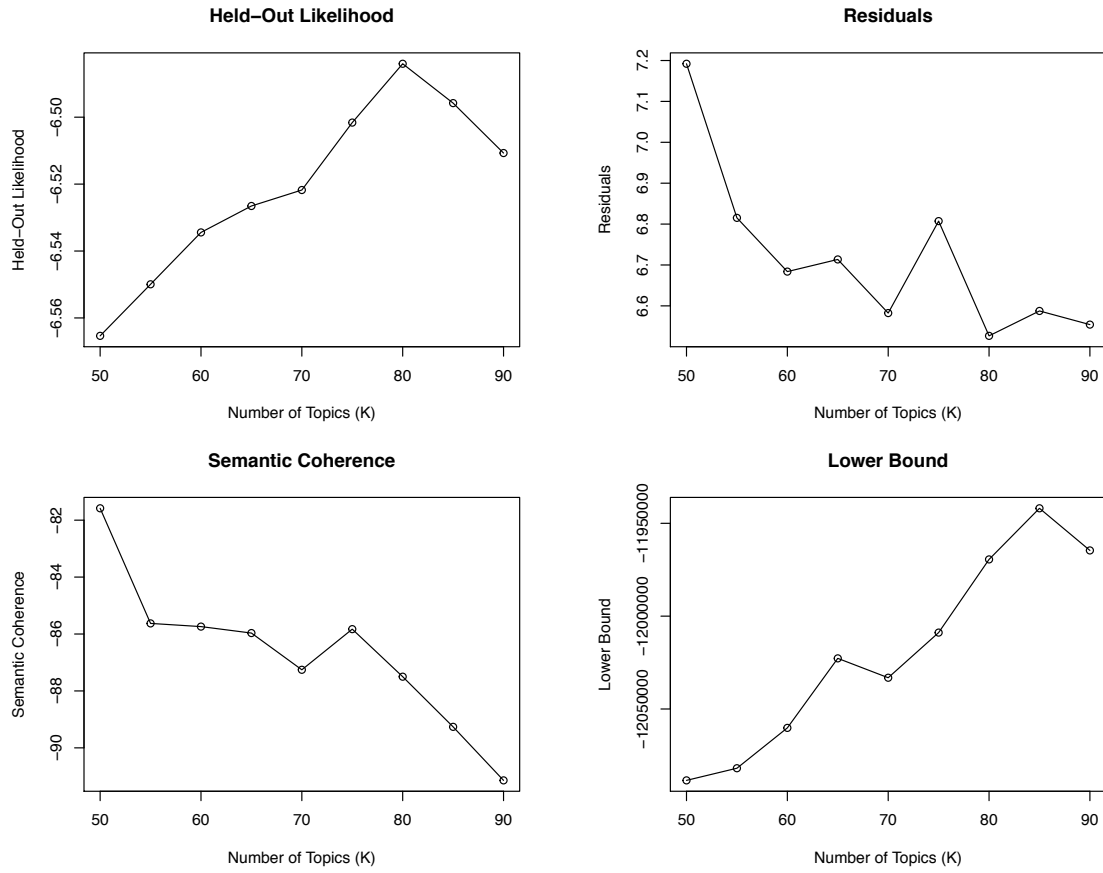


FIGURE C.1: Evaluation of K 's for STM part 1

Diagnostic Values by Number of Topics

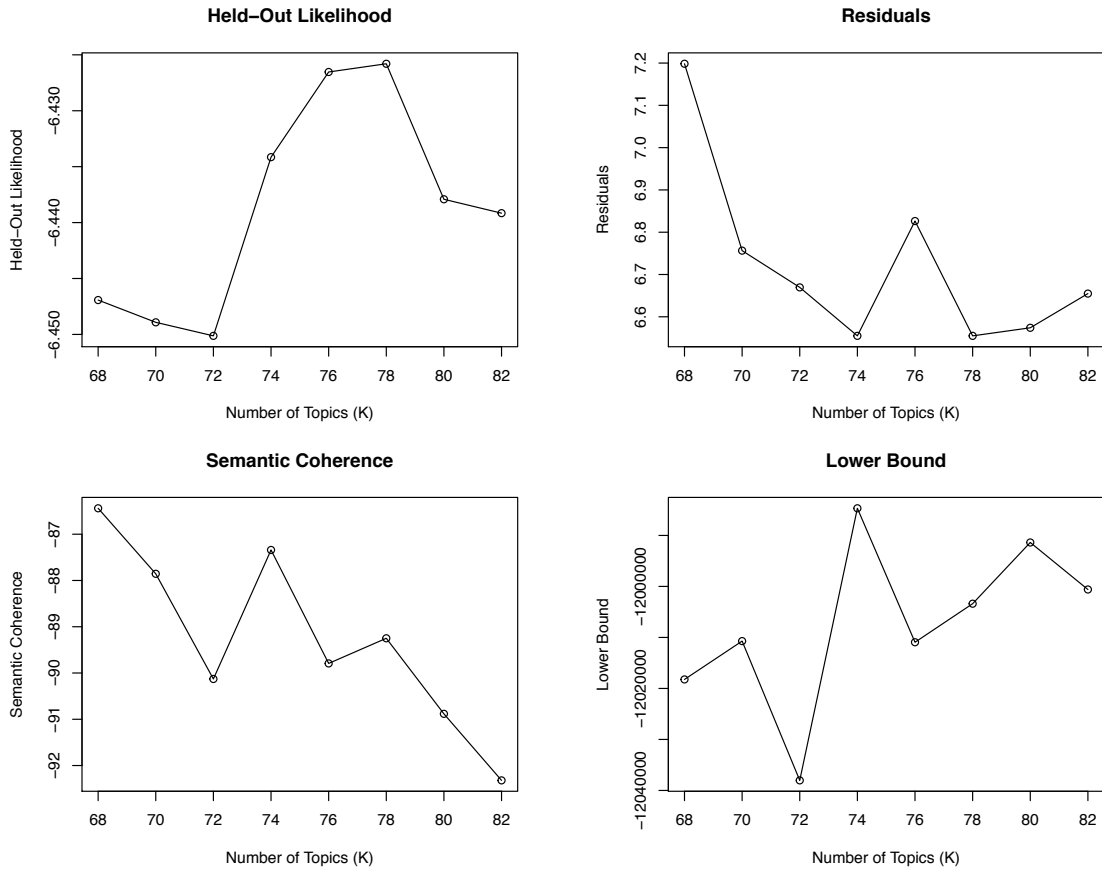


FIGURE C.2: Evaluation of K 's for STM part 2

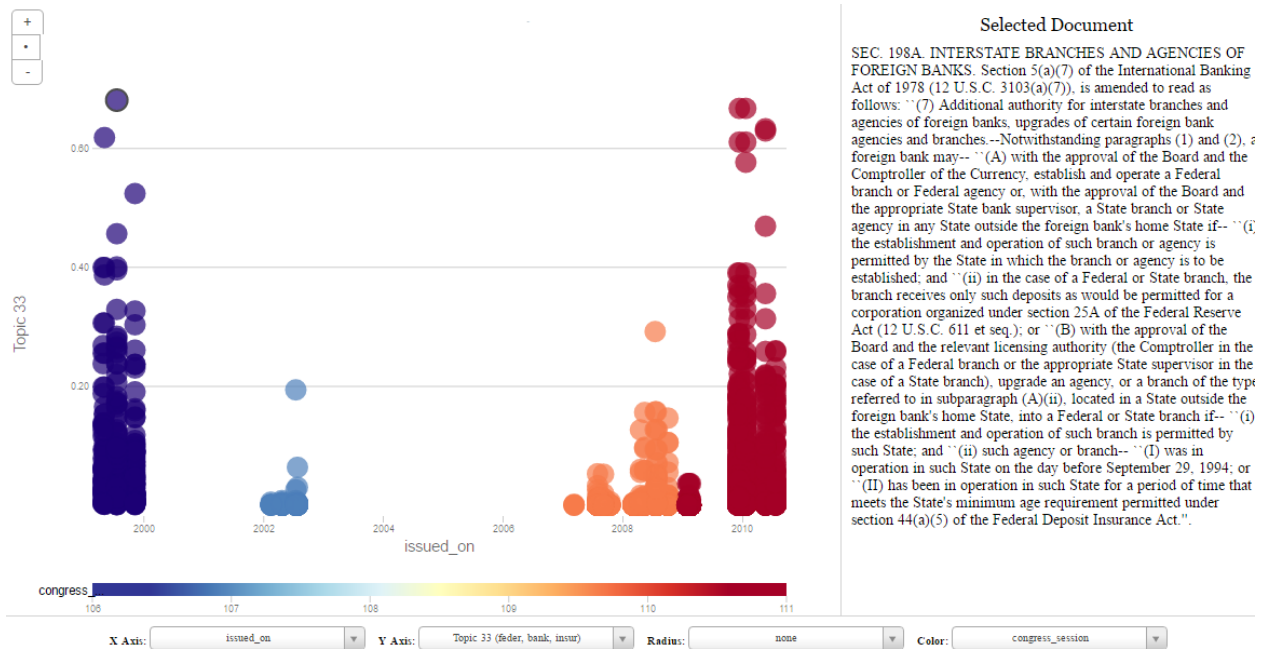


FIGURE C.3: Banking Topic Over All Bills

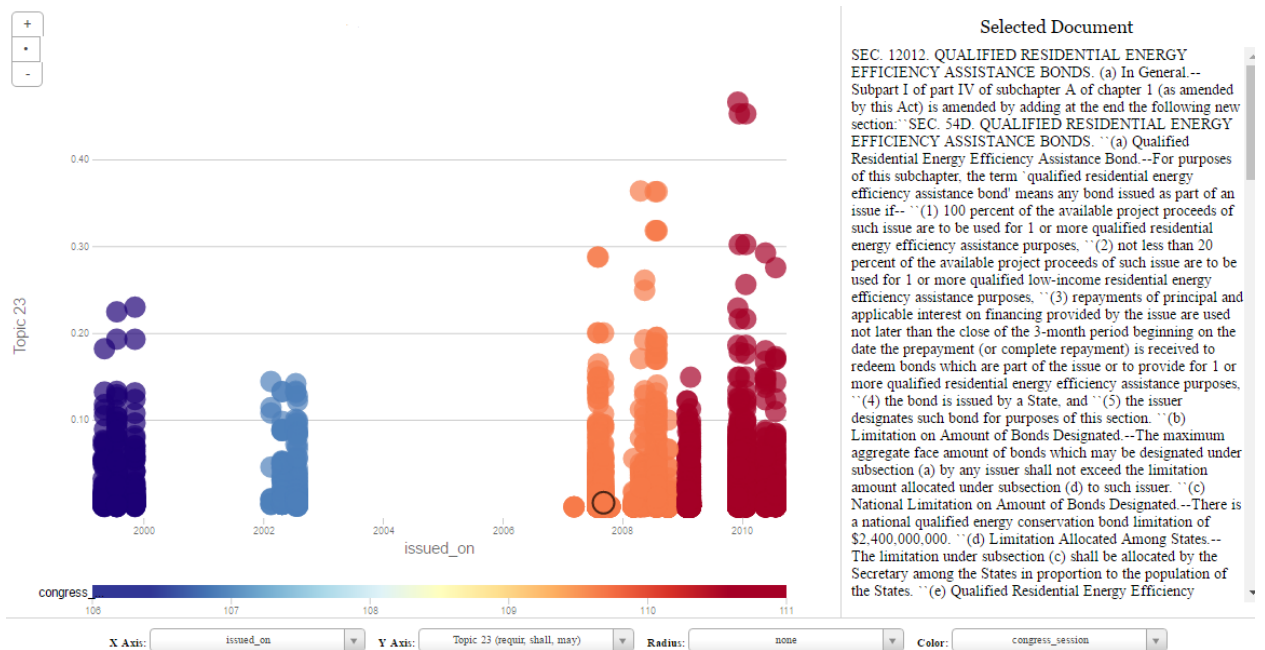


FIGURE C.4: Delegation Topic Over All Bills

Topic Effects on Reuse Instances

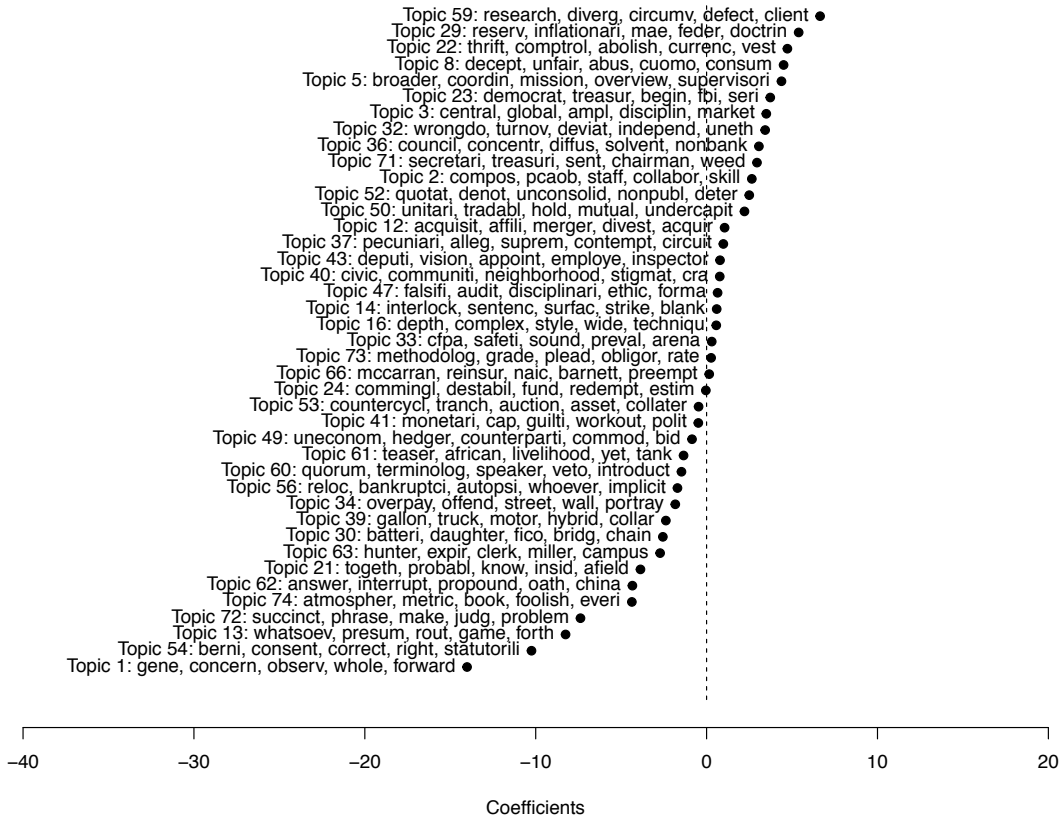


FIGURE C.5: LASSO of Topics on Text Reuse

Topic Effects on Reuse

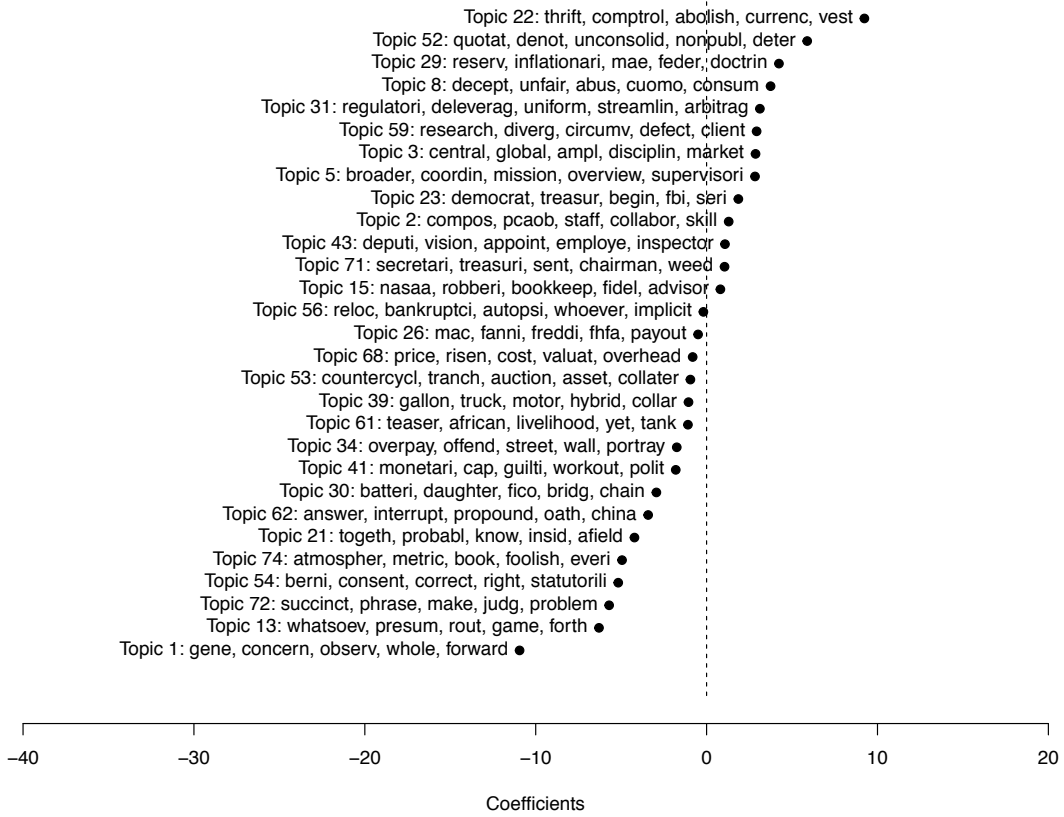


FIGURE C.6: LASSO of Topics on Text Counts

Bibliography

- Abelson, Donald E. 2009. *Do Think Tanks Matter?: Assessing the Impact of Public Policy Institutes*. Montreal: McGill-Queens University Press.
- Abramowitz, Alan I., and Kyle L. Saunders. 2008. "Is Polarization a Myth?" *Journal of Politics* 70 (2): 542-55.
- Adler, E. Scott, and John D. Wilkerson. 2013. *Congress and the Politics of Problem Solving*. Cambridge: Cambridge University Press.
- Adler, E. Scott. 2002. "New Issues, New Members: Committee Composition and the Transformation of Issue Agendas on the House Banking and Public Works Committees." *Policy Dynamics*, 230-49.
- Aldrich, John H. 2011. *Why Parties? A Second Look*. Chicago: University of Chicago Press.
- Anderson, William J., Jane M. Box-Steffensmeier, and Valeria Sinclair Chapman. 2003. "The Keys to Legislative Success in the U.S. House of Representatives." *Legislative Studies Quarterly* 28 (3): 357-86.

- Ansolabehere, Stephen, James M. Snyder, and Charles Stewart III. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45 (1): 136-59.
- Ball, Molly. 2013. "The Fall of the Heritage Foundation and the Death of Republican Ideas." *The Atlantic*, September 25. <https://www.theatlantic.com/politics/archive/2013/09/the-fall-of-the-heritage-foundation-and-the-death-of-republican-ideas/279955/>.
- Bartels, Larry M. 2005. "Homer Gets a Tax Cut: Inequality and Public Policy in the American Mind." *Perspectives on Politics* 3 (1): 15-31.
- Baumgartner, Frank R., and Bryan D. Jones. 2002. *Policy Dynamics*. Chicago: University of Chicago Press.
- Baumgartner, Frank R., and Bryan D. Jones. 2010. *Agendas and Instability in American Politics*. Chicago: University of Chicago Press.
- Baumgartner, Frank, Jeffrey M. Berry, Marie Hojnacki, David C. Kimball, and Beth L. Leech. 2009. *Lobbying and Policy Change: Who Wins, Who Loses, and Why*. Chicago: University of Chicago Press.
- Berry, Frances Stokes, and William D. Berry. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis." *American Political Science Review* 84 (2): 395-415.
- Beyer, Anne, Daniel A. Cohen, Thomas Z. Lys, and Beverly R. Walther. 2010. "The Financial Reporting Environment: Review of the Recent Literature." *Journal of Accounting and Economics* 50 (2): 296-343.
- Binder, Sarah. 2003. *Stalemate: Causes and Consequences of Legislative Gridlock*. Washington, D.C.: Brookings Institution Press.

- Blei, David M. 2012. "Probabilistic Topic Models." *Communications of the ACM* 55 (4): 77-84.
- Blei, David M., and John D. Lafferty. 2006. "Dynamic Topic Models." In *Proceedings of the 23 International Conference on Machine Learning*, pp. 113-120. ACM.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993-1022.
- Bloomfield, Louis. 2012. "WCOPYFIND." *The Plagiarism Resource Site*. <http://plagiarism.bloomfieldmedia.com/wordpress/software/wcopyfind/>.
- Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58 (2): 367-86.
- Bonica, Adam, Chen Jowei, and Tim Johnson. 2015. "Senate Gate-Keeping, Presidential Staffing of 'Inferior Offices,' and the Ideological Composition of Appointments to the Public Bureaucracy." *Quarterly Journal of Political Science* 10 (1): 5-40.
- Checkel, Jeffrey T. 1998. "The Constructive Turn in International Relations Theory." *World Politics* 50 (2): 324-48.
- Clinton, Joshua D., and John S. Lapinski. 2006. "Measuring Legislative Accomplishment, 1877-1994." *American Journal of Political Science* 50(1): 232-49.
- Clinton, Joshua D., Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98 (2): 355-70.
- Clinton, William J. 1999. "Statement on Signing the Gramm-Leach-Bliley Act." November 12. Gerhard Peters and John T. Woolley, *The American Presidency Project*. <http://www.presidency.ucsb.edu/ws/?pid=56922>.

- Cohen, Jeffrey E. 1986. "The Dynamics of the 'Revolving Door' on the FCC." *American Journal of Political Science* 30 (4): 689-708.
- Cox, Gary W., and Matthew D. McCubbins. 1993. *Legislative Leviathan: Party Government in the House*. Berkeley: University of California Press.
- Cox, Gary W., and Matthew D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the US House of Representatives*. Cambridge: Cambridge University Press.
- Crawford, Corinne. 2011. "The Repeal of the Glass-Steagall Act and the Current Financial Crisis." *Journal of Business and Economics Research* 9 (1): 127.
- Curry, James M. 2015. *Legislating in the Dark: Information and Power in the House of Representatives*. Chicago: University of Chicago Press.
- DeGregorio, Christine A. 1999. *Networks of Champions: Leadership, Access, and Advocacy in the U.S. House of Representatives*. Ann Arbor, MI: University of Michigan Press.
- DeGregorio, Christine. 1992. "Leadership Approaches in Congressional Committee Hearings." *Western Political Quarterly* 45 (4): 971-83.
- Denzau, Arthur T., and Michael C. Munger. 1986. "Legislators and Interest Groups: How Unorganized Interests Get Represented." *American Political Science Review* 80 (1): 89-106.
- Diermeier, Daniel, and Timothy J. Feddersen. 2000. "Information and Congressional Hearings." *American Journal of Political Science* 44 (1): 51-65.
- Edwards, Lee. 1997. *The Power of Ideas: The Heritage Foundation at 25 Years*. Ottawa, IL: Jameson Books.

- Elgie, Robert, and Iain McMenamin. 2005. "Credible Commitment, Political Uncertainty or Policy Complexity? Explaining Variations in the Independence of Non-Majoritarian Institutions in France." *British Journal of Political Science* 35 (3): 531-48.
- Epstein, David, and Sharyn O'Halloran. 1995. "A Theory of Strategic Oversight: Congress, Lobbyists, and the Bureaucracy." *Journal of Law, Economics, and Organization* 11 (2): 227-55.
- Epstein, David, and Sharyn O'Halloran. 1999. *Delegating Powers: A Transaction Cost Politics Approach to Policy Making under Separate Powers*. Cambridge: Cambridge University Press.
- Esterling, Kevin. 2004. *The Political Economy of Expertise: Information and Efficiency in American National Politics*. Ann Arbor, MI: University of Michigan Press, 2004.
- Esterling, Kevin M. 2007. "Buying Expertise: Campaign Contributions and Attention to Policy Analysis in Congressional Committees." *American Political Science Review* 101 (1): 93-109.
- Fauconnier, Gilles, and Mark Turner. 2008. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.
- Fenno, Richard F. 1966. *The Power of the Purse: Appropriations Politics in Congress*. New York: Little Brown.
- Fiorina, Morris P., Samuel A. Abrams, and Jeremy C. Pope. 2008. "Polarization in the American Public: Misconceptions and Misreadings." *Journal of Politics* 70 (2): 556-60.
- Fleisher, Richard, and John R. Bond. 2004. "The Shrinking Middle in the US Congress." *British Journal of Political Science* 34 (3): 429-51.

- Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. 2008. "A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models." *Annals of Applied Statistics* 2 (4): 1360-1383.
- Gerrish, Sean M., and David M. Blei. 2010. "A Language-based Approach to Measuring Scholarly Impact." In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 375-82.
- Gilens, Martin. 2001. "Political Ignorance and Collective Policy Preferences." *American Political Science Review* 95 (2): 379-96.
- Grimmer, Justin. 2010. "A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases." *Political Analysis* 18 (1): 1-35.
- Grimmer, Justin. 2013. *Representational Style in Congress: What Legislators Say and why it Matters*. Cambridge: Cambridge University Press.
- Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267-97.
- Groseclose, Tim, and Jeffrey Milyo. 2005. "A Measure of Media Bias." *Quarterly Journal of Economics* 120 (4): 1191-1237.
- Groseclose, Tim, Steven D. Levitt, and James M. Snyder Jr. 1999. "Comparing Interest Group Scores Across Time and Chambers: Adjusted ADA scores for the US Congress." *American Political Science Review* 93 (91): 33-50.
- Grossmann, Matt. 2012. *The Not-So-Special Interests: Interest Groups, Public Representation, and American Governance*. Palo Alto: Stanford University Press.

- Haeder, Simon F., and Susan Webb Yackee. (2015). "Influence and the Administrative Process: Lobbying the US President's Office of Management and Budget." *American Political Science Review* 109 (3): 507-22.
- Hall, Richard L. 1996. *Participation in Congress*. New Haven, CT: Yale University Press.
- Hall, Richard L., and Frank W. Wayman. 1990. "Buying Time: Moneyed Interests and the Mobilization of Bias in Congressional Committees." *American Political Science Review* 84 (3): 797-820.
- Hansen, John Mark. 1991. *Gaining Access: Congress and the Farm Lobby, 1919-1981*. Chicago: University of Chicago Press.
- Harbridge, Laurel. 2015. *Is Bipartisanship Dead?* Cambridge: Cambridge University Press.
- Hellebust, Lynn. 2001. *Think Tank Directory: A Guide to Nonprofit Public Policy Research Organizations*. Washington, DC: Government Research Service.
- Kiewiet, D. Roderick and Mathew McCubbins. 1991. *The Logic of Delegation*. Chicago: University of Chicago Press.
- Kingdon, John. 1984. *Agendas, Alternative, and Public Policies*. New York: Harper Collins
- Kingdon, John W. 1989. *Congressmen's Voting Decisions*. Ann Arbor: University of Michigan Press.
- Koger, Gregory, Seth Masket, and Hans Noel. 2009. "Partisan Webs: Information Exchange and Party Networks." *British Journal of Political Science* 39 (3): 633-53.
- Krehbiel, Keith. 1998. *Pivotal Politics: A Theory of U.S. Lawmaking*. Chicago: University of Chicago Press.

- Krutz, Glen S. 2005. "Issues and Institutions: Winnowing in the U.S. Congress." *American Journal of Political Science* 49 (2): 313-26.
- LaForge, William N. 2010. *Testifying Before Congress: A Practical Guide to Preparing and Delivering Testimony Before Congress and Congressional Hearings for Agencies, Associations, Corporations, Military, NGOs, and State and Local Officials*. Alexandria, VA: TheCapital.Net.
- Lauderdale, Benjamin E., and Tom S. Clark. 2014. "Scaling Politically Meaningful Dimensions Using Texts and Votes." *American Journal of Political Science* 58 (3): 754-71.
- Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97 (2): 311-31.
- Lerner, Joshua Y. 2017. "Getting the Message Across: Evaluating Think Tank Influence in Congress." Working Paper. https://sites.duke.edu/lerner/files/2015/04/Think_Tanks_Submission.pdf
- Lowe Will. 2008. "Understanding Wordscores." *Political Analysis* 16 (4): 356-71.
- Lupia, Arthur, and Mathew D. McCubbins. 1998. *The Democratic Dilemma. Can Citizens Learn What They Need to Know?* Cambridge: Cambridge University Press.
- MacKay, David J. C. 2003. *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press.
- Maltzman, Forrest, and Charles R. Shipan. 2008. "Change, Continuity, and the Evolution of the Law." *American Journal of Political Science* 52 (2): 252-67.
- Mayhew, David R. 1974. *Congress: The Electoral Connection*. New Haven, CT: Yale University Press.

- Mayhew, David R. 1991. *Divided We Govern: Party Control, Lawmaking, and Investigations, 1946-1990*. New Haven, CT: Yale University Press.
- McCarty, Nolan M., and Keith T. Poole. 1998. "An Empirical Spatial Model of Congressional Campaigns." *Political Analysis* 7 (1): 1-30.
- McCubbins, Mathew D., and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols Versus Fire Alarms." *American Journal of Political Science* 28 (1): 165-79.
- McCubbins, Mathew D., Roger G. Noll, and Barry R. Weingast. 1987. "Administrative Procedure as Instruments of Political Control." *Journal of Law, Economics, and Organization* 3 (2): 243-77.
- McCubbins, Matthew D., Roger G. Noll, and Barry R. Weingast. 1989. "Structure and Process, Politics and Policy: Administrative Arrangements and the Political Control of Agencies." *Virginia Law Review* 75 (2): 431-82.
- McGann, James G., ed. 2007. *Think Tanks and Policy Advice in the US: Academics, Advisors and Advocates*. New York: Routledge.
- McKay, Amy. 2008. "A Simple Way of Estimating Interest Group Ideology." *Public Choice* 136 (1-2): 69-86.
- Medvetz, Thomas. 2012. *Think Tanks in America*. Chicago: University of Chicago Press.
- Meltzer, Allan H. 2009. "Reflections on the Financial Crisis." *Cato Journal* 2: 25.
- Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.
- Piotroski, Joseph D., and Suraj Srinivasan. 2008. "Regulation and Bonding: The Sarbanes-

- Oxley Act and the Flow of International Listings.” *Journal of Accounting Research* 46 (2): 383-425.
- Poole, Keith T., and Howard Rosenthal. 1985. “A Spatial Model for Legislative Roll Call Analysis.” *American Journal of Political Science* 29 (2): 357-84.
- Poole, Keith T., and Howard Rosenthal. 1991. “Patterns of Congressional Voting.” *American Journal of Political Science* 35 (1): 228-78.
- Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.
- Posner, Richard A. 2009. *A Failure of Capitalism: The Crisis of '08 and the Descent into Depression*. Cambridge, MA: Harvard University Press.
- Rich, Andrew. 2005. *Think Tanks, Public Policy, and the Politics of Expertise*. Cambridge: Cambridge University Press.
- Roberts, Margaret E., Brandon M. Stewart, and Edoardo M. Airoidi. 2016. “A Topic Model for Articles Experimentation in the Social Sciences.” *Journal of the American Statistical Association*. 111 (515): 988-1003.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. “Structural Topic Models for Open-Ended Survey Responses.” *American Journal of Political Science* 58 (4): 1064-1082.
- Shepsle, Kenneth A. & Barry R. Weingast. 1987. “The Institutional Foundations of Committee Power.” *American Political Science Review* 81(1): 85-104.
- Shipan, Charles R., and Craig Volden. 2008. “The Mechanisms of Policy Diffusion.” *Amer-*

- ican Journal of Political Science* 52 (4): 840-57.
- Shor, Boris, and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105 (3): 530-51.
- Slapin, Jonathan B., and Sven Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52 (3): 705-22.
- Stone, Diane. 1996. *Capturing the Political Imagination: Think Tanks and the Policy Process*. London: Frank Cass.
- Tausanovitch, Chris, and Chris Warshaw. 2012. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." *Journal of Politics* 75 (2): 330-42.
- Tesler, Michael. 2012. "The Spillover of Racialization into Health Care: How President Obama Polarized Public Opinion by Racial Attitudes and Race." *American Journal of Political Science* 56 (3): 690-704.
- Treasury, U. S. Department of. 2009. *Financial Regulatory Reform: A New Foundation—Rebuilding Financial Supervision and Regulation*. https://www.treasury.gov/initiatives/Documents/FinalReport_web.pdf
- Troy, Tevi. 2012. "Devaluing the Think Tank." *National Affairs* (Winter). <http://www.nationalaffairs.com/publications/detail/devaluing-the-think-tank>.
- Turner, Mark. 2001. *Cognitive Dimensions of Social Science*. New York: Oxford University Press.
- Turner, Mark. 2014. *The Origin of Ideas: Blending, Creativity, and the Human Spark*. New York: Oxford University Press.

- Volden, Craig, Alan E. Wiseman & Dana E. Wittmer. 2013. "When Are Women More Effective Lawmakers Than Men." *American Journal of Political Science* 57 (2): 326-41.
- Volden, Craig, Michael M. Ting, and Daniel P. Carpenter. 2008. "A Formal Model of Learning and Policy Diffusion." *American Political Science Review* 102 (3): 319-32.
- Walker, Jack, Jr. 1983. "The Origins and Maintenance of Interest Groups in America." *American Political Science Review* 77 (2): 390-406.
- Weaver, R. Kent. 1989. "The Changing World of Think Tanks." *PS: Political Science and Politics* 22 (3): 563-78.
- Weidenbaum, Murray L. 2011. *The Competition of Ideas: The World of the Washington Think Tanks*. New Brunswick, NJ: Transaction Publishers.
- Wilkerson, John, Nick Stramp, and David Smith. 2015. "Tracing the Flow of Policy Ideas in Legislation: A Text Reuse Approach." *American Journal of Political Science* 59 (4): 943-56.

Biography

Joshua Yoshio Lerner was born November 18th, 1989 in Northampton Massachusetts. He grew up in Rockville, MD, graduating with honors from Rockville High School in 2008. He attended the University of Chicago where he obtained his Bachelors of Arts in Political Science in 2012. Following this, he enrolled at Duke University in the PhD program in Political Science, obtaining his Masters in 2014, and finishing this dissertation in 2017.

His primary research interests are the study of Congress, Text-as-Data, Political Ideology, Inter-branch relations, Machine Learning, and Causal Inference methods. His dissertation deals with applications of text-as-data approaches to the study of idea transmission in Congress, looking specifically at the amount of text-borrowing that occurs between testimony during committee hearings, and the language of final bills.

Starting in July, Josh will be a Postdoctoral Research Fellow at the Pritzker School of Law, Northwestern University.