

© 2016 Michael Zachary Gill

All rights reserved.

ProQuest Number: 10663789

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10663789

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

*Dissertation Advisor:*  
**Professor Gary King**

*Author:*  
**Michael Zachary Gill**

## **Empirical Essays on Secrecy and Security in the United States**

### **Abstract**

This dissertation analyzes longstanding issues in U.S. foreign policy and political economy with novel data and research methods. Chapter 1 asks: to what extent do “surprise” shifts in the international security environment help individual firms in the defense economy? This paper exploits the timing of three major events—the terrorist attacks of 9/11, the announcement of the Iraq troop surge, and the death of Osama Bin Laden—to assess how firms financially respond to shocks in the demand for defense. Utilizing financial market data for a set of “exposed” defense firms, event studies are performed via the estimation of Bayesian structural time series models. Results of the analysis demonstrate there is considerable heterogeneity in firm response to major events—and the estimation strategy notably outperforms competing estimators, such as the popular synthetic control method.

Chapter 2 (co-authored with Arthur Spirling) studies the role of information dissemination and communication at the U.S. Department of State from a unique empirical perspective. We analyze over 163,958 United States diplomatic cables to speak to several aspects of contemporary international relations theory. We show that diplomatic secrecy consists of at least two distinct “dimensions”: substantive and procedural. The former deals with secrets per se relating to specific political issue areas that would actively damage U.S. interests, especially in terms of revealing the resolve or capabilities. Procedural secrecy deals with the diplomatic norm of confidentiality in meetings, regardless of the substantive content of any single cable. We relate these two dimensions of diplomacy to concepts of secrecy in the theoretical literature, and demonstrate that both play an important role in the bureaucratic behavior of the U.S. Foreign Service.

Chapter 3 (co-authored with Michael Egedal and Martin Rotemberg) analyzes how the behavior of the Federal Open Market Committee changed after the statutory enforcement of transparency laws in 1993; to do so, we present new techniques to describe how language use changes over time. For a family of widely used vector space metrics, we demonstrate how to decompose aggregate changes into each individual dimension's contribution—such as a particular word or document's influence. The approach can be generalized to account for associations between document dimensions (such as word definitions or meanings). Using various documents released by the Federal Reserve from 1976 to 2007, covering both years in which the FOMC knew its deliberations would eventually be made public, and years in which it believed no records were kept, we find that FOMC deliberations became more similar to the always-public press releases in the transparency regime.

# Contents

Abstract . . . . .	iii
Acknowledgments . . . . .	xii
<b>Introduction</b>	<b>1</b>
<b>1 The Economic Benefits of Conflict? Estimating Defense Firm Responses to Major Events in U.S. Foreign Policy</b>	<b>4</b>
1.1 Introduction . . . . .	4
1.2 The Disconnect Between Firm Valuations, Defense Procurement, and War . . . . .	9
1.3 Empirical Strategy . . . . .	11
1.3.1 Notation and Framework . . . . .	11
1.3.2 Statistical Model: Bayesian Structural Time Series with Synthetic Controls . . . . .	13
1.3.3 Selection of Control Variables . . . . .	19
1.3.4 Study Windows and External Validity . . . . .	24
1.3.5 Data Sources . . . . .	26
1.4 Results: Firm Financial Responses to 9/11 Attacks, the “Troop Surge” in Iraq, and Death of Osama Bin Laden . . . . .	27
1.4.1 Firm-level Results . . . . .	27
1.4.2 Cross-Sectional Heterogeneity . . . . .	30
1.4.3 In-Sample Performance of Competing Statistical Approaches . . . . .	37
1.5 Monte Carlo Studies to Assess Model Performance . . . . .	40
1.5.1 Simulation Study 1: Random Correlations between Exposed Series and Controls . . . . .	41
1.5.2 Simulation Study 2: A Multi-factor Model with Random Dependence on Principal Components . . . . .	47
1.6 Discussion . . . . .	52

<b>2</b>	<b>Dimensions of Diplomacy: Understanding Private Information in U.S. Foreign Policy Using the WikiLeaks Cable Disclosure</b>	<b>58</b>
2.1	Introduction . . . . .	58
2.2	The Study of Private Information Disclosure . . . . .	62
2.2.1	Implications and Expectations . . . . .	65
2.3	Data . . . . .	67
2.4	Methods . . . . .	72
2.4.1	Substantive Secrecy . . . . .	72
2.4.2	Procedural Secrecy . . . . .	73
2.5	Results . . . . .	78
2.5.1	Substantive Secrecy: High vs. Low Politics . . . . .	78
2.5.2	Matched Sample Results: Procedural Secrecy . . . . .	80
2.5.3	Share of Secrecy: Substance vs. Procedure . . . . .	83
2.6	Discussion . . . . .	85
<b>3</b>	<b>A Causal Text Analysis of How Federal Reserve Discussions Respond to Increased Transparency.</b>	<b>88</b>
3.1	Introduction . . . . .	88
3.2	Historical Background . . . . .	95
3.2.1	Data availability . . . . .	95
3.2.2	Data . . . . .	97
3.3	Methodology . . . . .	97
3.3.1	Notation . . . . .	97
3.3.2	Similarity Metric Axioms . . . . .	98
3.3.3	Cosine Similarity and Generalized Cosine Similarity . . . . .	99
3.3.4	Growth in Cosine Similarity . . . . .	101
3.3.5	Constructing the Term-Relationship Weight Matrix . . . . .	104
3.4	Results . . . . .	105
3.4.1	The Evolution of Language after Transparency . . . . .	105
3.5	Conclusion . . . . .	116
	<b>Appendices</b>	<b>118</b>
A	Appendix to Chapter 1 . . . . .	119
A.1	Supplementary Analyses . . . . .	119
A.2	Product Responsiveness to War Escalation . . . . .	123
A.3	Data Sources and Cleaning . . . . .	124
A.4	Summary of Covariates Included in Final BSTS Models . . . . .	131
A.5	Market Competition Across Defense Products . . . . .	132

A.6	Firm-Level Results Across Competing Estimators . . . . .	133
A.7	Supplemental Tables and Figures . . . . .	136
B	Appendix to Chapter 2 . . . . .	140
B.1	Possible Data Limitations . . . . .	140
B.2	How Cables are Written and Classified . . . . .	141
B.3	Additional Details on Lasso and Random Forest . . . . .	143
B.4	Cable Tags . . . . .	145
B.5	Average Cable Restrictiveness by Embassy . . . . .	146
B.6	Reduction in Imbalance from Matching . . . . .	147
B.7	Counts by Embassy in Matched Sample . . . . .	148
B.8	Exact Matching Algorithm . . . . .	149
B.9	Probabilistic Topic Model . . . . .	150
C	Appendix to Chapter 3 . . . . .	151
C.1	Derivations of Growth in Similarity Measures . . . . .	151
C.2	Topic modeling with a dictionary . . . . .	153
C.3	A bound on bias in similarity when trimming vectors . . . . .	156
C.4	Effects of Sample Size . . . . .	158
C.5	Theoretical Model . . . . .	159

**References**

**164**

## List of Tables

1.1	Correlations of Daily Prices and Returns for Prominent Firms (Two Years of Trading Data Prior to 9/11)	42
1.2	BSTS Performance in Principal Components Experiment (Study 2)	50
3.1	Effect of Transparency on Document Similarity	108
A.1	Two-Digit Product Codes in DoD Procurement Data	126
A.2	BSTS Variable Inclusion Probabilities Across Events and Models	131



## List of Figures

1.1	Differing Trajectories of Observed Firm-Level Financial Returns After Major Events . . . . .	12
1.2	Firm-Level Responses to 9/11 Attacks . . . . .	31
1.3	Firm-Level Responses to News of Iraq Troop Surge . . . . .	32
1.4	Firm-Level Responses to Osama Bin Laden's Death . . . . .	33
1.5	How Firm-Level Abnormal Returns Map to Quantiles of DoD Revenue Dependence . . . . .	35
1.6	Abnormal Returns versus Company Characteristics (High-Earning Subset of Firms) . . . . .	36
1.7	Relative Fit of Counterfactual Estimates in Pre-Event Window (Scaled to Baseline BSTS Model) . . . . .	39
1.8	Autocorrelation of Stocks in Sample (Example: Northrop Grumman Daily Share Price) . . . . .	43
1.9	BSTS Performance versus Variable Inclusion Rates with Data from Random Covariance Matrices (Study 1) . . . . .	46
1.10	Performance of Competing Estimators (Study 2) . . . . .	51
2.1	Number of cables per month in Sample, 1966–2010 . . . . .	68
2.2	Distribution of tag counts in sample . . . . .	70
2.3	Conditional probability of U.S. State Department subject TAG co-occurrences in the post-2005 sample ( $n = 163,958$ ) . . . . .	71
2.4	Substantive content as a predictor of secrecy status in Full Sample (Substantive Secrecy Analysis) . . . . .	79
2.5	Words Most Strongly Predictive of Secrecy in Matched Sample (Procedural Secrecy Analysis) . . . . .	82
2.6	Classification Performance of Models that Incorporate Words, Tags, and Embassy-Level Information . . . . .	84

3.1	Changing Naming Conventions of Federal Reserve documents from 1967–2007 . . . . .	95
3.2	Cosine similarity of FOMC transcripts and corresponding public summaries from 1976–2007 . . . . .	106
3.3	Approximate and observed growth for cosine and generalized cosine similarity . . . . .	109
3.4	Extended cosine similarity of FOMC transcripts and corresponding public summaries from 1976–2007 . . . . .	110
3.5	Decomposition of similarity growth into public and private document contributions . . . . .	111
3.6	Permutation tests of the change in cosine similarity . . . . .	113
3.7	Permutation tests of the change in generalized cosine similarity . . . . .	114
3.8	Cumulative contribution of word’s similarity growth . . . . .	115
3.9	Relative contributions by individual words to growth in cosine similarity and generalized cosine similarity . . . . .	117
A.10	How Firm-level Product Sales Map to Causal Effects . . . . .	121
A.11	Proportion of Each Firm’s Total Obligations Awarded “Non-competitively” (FY2000 to September 10, 2001) . . . . .	122
A.12	How Product Purchases Grow Alongside Troop Levels (All DoD Obligations Prior to Iraq Surge Announcement) . . . . .	123
A.13	Daily Outlays to Top-Grossing PSCs (FY2000 onwards) . . . . .	127
A.14	Monthly U.S. Troop Levels in Afghanistan and Iraq . . . . .	129
A.15	R Script for Matching Firm Names Across Databases . . . . .	130
A.16	Higher-Earning Defense Products Less Likely to be “Competitively” Awarded (FY2000 to September 10, 2001) . . . . .	132
A.17	Relations Between Firm-Level Causal Estimates Across Competing Estimators and Events . . . . .	135
A.18	Proportion of Products Sold to DoD by Firm (FY2009 to FY2011) . . . . .	136
A.19	Firm and Sector Returns vs. Benchmark Indices . . . . .	137
A.20	Monthly DoD Outlays (2000–2014) . . . . .	138
A.21	Number of Firms Receiving DoD Outlays in each Month . . . . .	139
B.22	State Department “Subject TAGS” in sample and meanings . . . . .	145
B.23	Frequency of restricted versus unrestricted cables by place of origin, for all cables post-2005. . . . .	146
B.24	Reduction in Subject Imbalance From Exact Matching . . . . .	147
B.25	Counts of cables by embassy in the matched sample. . . . .	148
B.26	Outline of Exact Matching Algorithm . . . . .	149

C.27 RTM topics in the ODE . . . . .	157
C.28 Close to no association between document length and similarity in the pretreatment period . . . . .	159
C.29 Simulated change in similarity using documents generated from observed word proportions . . . . .	160

## Acknowledgments

Over six years, I've learned that it *is*, in fact, possible to write a dissertation; but I've also learned it is hard to write a good one, and nearly impossible to write anything of consequence without support from others. First, I want to thank my advisors for their loyalty, compassion, and guidance. Gary King, Dustin Tingley, Jeff Frieden, and Ken Shepsle are the scholars I aim to be. They have taught me that social science can be both technical and immensely creative, that complexity often comes at the cost of clarity, that research begins with asking a great question, and that the work we do can change lives. To be able to say that "Gary, Dustin, Jeff, and Ken taught me political science" is among the greatest privileges one can have in our discipline, and one I certainly feel I don't deserve. I hope that as the years go on I will make them proud.

Arthur Spirling—while not technically an advisor—has been a mentor and central part of my life over the last several years. He is a dream co-author, and an even better friend. He represents the future of our discipline. Some of my favorite moments from graduate school were spent brainstorming in his office, scribbling half-formed (if not quarter-formed) ideas on his whiteboard (with Arthur patiently nodding along). Through our friendship I have also gotten to know Amy—no doubt, the better half of the Spirling power couple—and their beautiful daughter, Elizabeth. I am genuinely honored to have the chance to continue my friendship and work with Arthur and Amy in a new city.

I met Andy Hall on one of the first days of graduate school orientation, and in the evening we went out to eat hamburgers. Since then, he and I have probably together (conservatively) consumed several hundred pounds of meat in different forms. We were roommates for two glorious years—an era marked by fantastic food, action movies, late night discussions about politics and statistics, and work that felt like play. He taught me to think creatively, that academic progress often comes by blending approaches and disciplines, and that it is often okay, if not preferred, to "take no notice" of the first attempt at something. He also invented the Cynaroni. I am indebted to Andy for more

than he could ever imagine, and so proud to call him one of my closest friends.

Mike Egesdal is my spirit animal in human form. I remember first meeting Mike back home in Hawaii as a small child, when my parents took me and my brother over to the Wichman's house at the top of Tantalus. The Egesdal boys were there playing Scrabble, if memory serves, sitting around a square table eating bread with Grandma's homemade guava jelly. They proceeded to beat the Gill boys handily. (It might have been my first game of Scrabble, but not my last.) Fast forward twenty-something years, and Mike is one of my best friends. He may be the kindest guy I've ever met, one of the most athletic, and certainly one of the smartest—a frustrating combination that, with time, I've learned to accept. Whenever I'm around him I smile, learn something new, and feel like I'm at home. In our most recent game of Scrabble, I am pleased to say I won: 321–319.

Countless others deserve thanks. My parents: for encouraging political discussion at dinner, allowing my female cat to be named Michelangelo, and expecting that Tommy and I form somewhat reasonable opinions. My brother, Tommy: for his unwavering friendship and early-life decision not to exploit his age and relative physical prowess to my disadvantage. My grandparents: Ana, Lois, Tom, and Sakari, for convincing me that education is important and that a life dedicated to the pursuit of truth can be mixed with public service. My ol' pappy: whose warming spirit continues to give me optimism for the future. The Cumming clan—Jack, Pam, Lacey, and Sean—for welcoming me into their homes, expanding my family, and providing love and support at all times, even amidst tragedy. Rambo and Koa: for being spiritual guides. Justin Abdallah, Muhammet Bas, Alexis Diamond, Adam Glynn, Alisa Hall, Connor Huff, Chris Lucas, Lindsay Majno, Sophie Merrifield, Jim Murray, Ellen Nicholson, Mike Pagliarini, Cooper Pickett, Martin Rotemberg, Tim Rutledge, Liza Ryan, Ben Schneer, Nina Schragar, Kunaal Sharma, Veronica Siverd, Alex VanderEls, Harlon Wheatley, and Fernando Yu: I love you all.

Some leave the best for last; in this case, I have. More than anyone else, Caitlin Cumming made this possible. She is my favorite person and the best part of my life.

To my family—my brother, parents, and grandparents.

## Introduction

This dissertation analyzes longstanding issues in U.S. foreign policy and political economy with novel data and research methods. Unifying themes across essays include how actors behave in response to surprise shifts in their political environments, and how political decision-makers communicate in public versus private settings. Across these cases, new statistical techniques are developed that contribute to literatures on performing causal inferences in event studies and the analysis of “text-as-data.”

Chapter 1 asks: to what extent do surprise shifts in the international security environment help or hinder individual firms in the defense economy? This paper exploits the timing of three major events—the terrorist attacks of 9/11, the announcement of the Iraq troop surge, and the death of Osama Bin Laden—to assess how individual firms financially respond to shocks in the demand for defense. Utilizing daily financial market data for a set of “exposed” defense firms (i.e., publicly-traded firms that received defense contracting dollars in prior fiscal years), financial event studies are performed through the estimation of Bayesian structural time series models. Counterfactual series are estimated using the observed behavior of unexposed firms and measures of market performance, and firm-level models are chosen through an embedded variable selection procedure. Results of the analysis demonstrate there is considerable heterogeneity in firm response to major events. Amongst prominent defense contractors, firm valuations routinely deviate from their (firm-level) expectations by billions of dollars after these events, and effects appear to last for both the short and long term. To inspect heterogeneous response, firm-level causal estimates are merged with official procurement records from the U.S.

Department of Defense and financial statistics disclosed to the Securities and Exchange Commission. Firms that respond most extremely to events tend to be more revenue dependent on the DoD. These findings suggest that events such as those studied not only lead to immediate shifts in beliefs over the demand for defense procurement, but firm-level effects can be largely explained by differences in their financial portfolios. Results from competing estimators, such as the popular “synthetic control” method are also discussed. Counterfactual series estimated via BSTS tend to outperform the traditional synthetic control estimates; however, BSTS models that incorporate traditional synthetic controls as potential covariates tend to perform the best overall. A range of simulation studies provides further evidence of these claims.

Chapter 2 (co-authored with Arthur Spirling) studies the role of information dissemination and communication at the U.S. Department of State from a unique empirical perspective. Noting that little systematic observational data exists regarding the contemporary private information available to foreign policy actors, we analyze over 163,958 United States diplomatic cables to speak to several aspects of contemporary international relations theory. We show that diplomatic secrecy consists of at least two distinct “dimensions”: substantive and procedural. The former deals with secrets per se relating to specific political issue areas that would actively damage U.S. interests, especially in terms of revealing the resolve or capabilities. Procedural secrecy deals with the diplomatic norm of confidentiality in meetings, regardless of the substantive content of any single cable. We relate these two dimensions of diplomacy to concepts of secrecy in the theoretical literature, and demonstrate that both play an important role in the bureaucratic behavior of the U.S. Foreign Service.

Chapter 3 (co-authored with Michael Egedal and Martin Rotemberg) analyzes how the behavior of the Federal Open Market Committee changed after the statutory enforcement of transparency laws in 1993. To do this, we develop techniques to describe how language use changes over time. For a family of widely used vector space metrics, we demonstrate



how to decompose aggregate changes into each individual dimension's contribution, such as a particular word's influence. Our approach can be generalized to account for associations between document dimensions (such as word definitions or meanings). Using various documents released by the Federal Reserve from 1976 to 2007, covering both years in which the FOMC knew its deliberations would eventually be made public, and years in which it believed no records were kept, we find that FOMC deliberations became more similar to the always-public press releases in the transparency regime. FOMC members shifted their comments towards popular economic subjects, such as "inflation" and "growth," and away from words that express personal opinions, like "think." In this setting, we demonstrate that the observed changes are not purely due to substitution across words with the same meaning, as the results are robust to accounting for semantic relations—a claim that would be difficult to make with pre-existing statistical techniques.

# Chapter 1

## The Economic Benefits of Conflict? Estimating Defense Firm Responses to Major Events in U.S. Foreign Policy<sup>1</sup>

### 1.1 Introduction

*“I’m not aware of any agency with the authority, responsibility or a process in place to coordinate all these interagency and commercial activities. The complexity of this system defies description.”*

— Lt. General John R. Vines

On the last trading day before 9/11, the Lockheed Martin Corporation—at the time, the largest defense contractor in the world, having received over \$15 Billion USD in prime awards from the U.S. Department of Defense (DoD) in FY2000—had a market capitalization of \$16.7 Billion USD dollars. On the first trading day after U.S. markets reopened, the price of Lockheed Martin stock increased over 20%, setting as record as the largest single-day return in the company’s history. Not all major “aerospace” firms seemed to benefit from the terrorist attacks, however. The Boeing Company—the second largest defense contractor in the world at the time, with over \$12 Billion USD dollars in contracts in FY2000—took a major hit after the terrorist attacks, dropping by over 17%

---

<sup>1</sup>Caitlin Cumming, Alexis Diamond, Mike Egedal, Martin Feldstein, Jeff Frieden, Joseph Foster, Andy Hall, Connor Huff, Gary King, Chris Lucas, Martin Rotemberg, Anne Sartori, Kunaal Sharma, Ken Shepsle, Arthur Spirling, Anton Streznev, Dustin Tingley, and Alex VanderEls are thanked for useful conversations, comments, and suggestions, as well as audience members at Harvard University, the Harvard-MIT-Yale Conference on Political Violence, and the Annual Meeting of the Midwest Political Science Association.

after markets reopened. This marked the largest single-day loss in Boeing's history. The two largest recipients of Department of Defense (DoD) dollars, the two largest aerospace firms in the world, had near-opposite reactions to the same event.

This paper asks: to what extent do major shifts in the global security environment help or hinder individual defense firms? To what extent do firm characteristics help explain heterogeneous responses to these events? It is widely known that defense sector's are tied to shifts government demand for goods and services (e.g., Ramey and Shapiro, 1999; Ramey, 2011). However, little to no scholarly research inspects how the financial wellbeing of individual firms actually reacts to shifts in the international security environment. This is despite the fact that in twenty-first century political life, there are near countless stakeholders in the global defense economy. Nearly all governments around the world rely on businesses for national security (Pattison, 2014; Singer, 2008). On an annual basis, arms expenditures alone represent more than 1 percent of global GDP (SIPRI). Since the Global War on Terror began, the DoD has been responsible for the employment of between 4 and 5 million workers per year, leading some to call the Defense Department "by far the largest and most complex business organization in the world" (Fox, 2011). More than half of all research and development funded by the American government goes to the defense sector (OMB, 2008).

While the politics of defense spending have always been deeply "political" (Clinton and Lewis, 2008; Mintz, 1989; Reich, 1972; Russett, 1976; Smith, 2009)—and financial interests, loosely defined, have a long history in security policy (Percy, 2007)—businesses are now more integrated in foreign affairs than ever before in human history (Garten, 1997; Jacobs and Page, 2005). The co-dependence between governments and firms has been alleged to complicate a government's ability to shift military strategy without having a direct influence on the livelihood of individual firms.<sup>2</sup> Former Secretary of Defense

---

<sup>2</sup>For example, major contractors like Northrop Grumman and Raytheon acknowledge the importance of both DoD contracts and their expectations over government demand in official company filings with the SEC. In Northrop Grumman's 2014 10-K form, item 1.A—which describes a company's major sources of

Robert Gates made note of this dynamic in a 2010 speech, where he spoke on the challenge of reducing the defense budget: “What it takes is the political will and willingness. . . to make hard choices—choices that will displease powerful people both inside the Pentagon, and out.”

This paper advances recent quantitative research on special interest behavior by analyzing how firm valuations are tied to the international security environment. Leveraging a new dataset on over 25,000,000 contract actions at the United States Department of Defense (DoD) with a total nominal value exceeding 4 trillion USD, firms are scaled according to the products they sell to the DoD and their overall revenue dependence to government contracts. For a sample of public firms on U.S. markets, precise measures of firm dependencies are measurable given financial disclosures to the U.S. Securities and Exchange Commission (SEC) in their quarterly and annual reports (i.e., 10-Q and 10-K documents). Firms are scaled by the degree to which individual products and services sold to the DoD (e.g., 10mm machine guns, homing missiles, statistical consulting services, vehicle repair).

The central contribution this paper makes is the documentation the defense sector’s considerable variation in financial response to major events foreign policy, and how that variation may be utilized to understand firm behavior and their political incentives more

---

financial risk—notes:

“We depend heavily on a single customer, the U.S. Government, for a substantial portion of our business. Changes in this customer’s priorities and spending could have a material adverse effect on our financial position, results of operations and/or cash flows. . . Significant delays or reductions in appropriations for our programs and U.S. Government funding more broadly may negatively impact our business and programs and could have a material adverse effect on our financial position, results of operations and/or cash flows. . . We use estimates when accounting for contracts. Contract cost growth or changes in estimated contract revenues and costs could affect our profitability and our overall financial position” (Northrop Grumman Corporation, 2014).

Raytheon acknowledges similar dynamics in its own 2014 10-K, noting: “We depend on the U.S. Government for a substantial portion of our business, and changes in government defense spending and priorities could have consequences on our financial position, results of operations and business. In 2014, U.S. Government sales, excluding foreign military sales, accounted for approximately 70% of our total net sales” (Raytheon Company, 2014).

broadly. Understanding whether, to what degree, and why individual firms benefit from major shifts in the international security environment is a challenge for the empirical research agenda on special interest politics (e.g., de Figueiredo and Richter, 2014). Many conventional research designs will struggle to provide meaningful insight into the realities firms face given the observational (i.e., non-experimental) nature of the data in question. Cross-sectional designs that pool firm characteristics (e.g., correlating contract awards with wartime characteristics and firm valuations) will not capture how firm welfare is tied to shifts in the security environment if, for example, there is a lag between government spending and publicly observable “news,” if firms themselves are able to directly influence shifts in the defense budget, or if firms may reasonably anticipate major events in international politics ahead of time (e.g., if firms have expectations over troop increases or decreases in a country given public speeches made by politicians). If firm-level outcomes, such as valuations or political behaviors, respond faster to “news” than government procurement or overall expenditures, the naïve comparison of firm traits with data such as government outlays or troop levels will be unlikely to provide unbiased causal inferences about how firm welfare is shaped by shifts in the security environment. Inspections into heterogenous effects at the firm level would similarly be complicated by these so-called “endogenous” dynamics.

To address such issues that impede straightforward analysis, a variety of techniques are employed in this study. First and foremost, quasi-experimental “event studies” are performed on the financial returns of publicly traded corporations following the unanticipated occurrence of three major events relevant to U.S. foreign policy: the terrorist attacks of September 11, 2001, President George W. Bush’s announcement in early 2007 of an eventual “troop surge” in Iraq, and the surprising killing of Osama bin Laden on May 2, 2011. Our analysis focuses on these three events in particular both because of their unquestionable relevance to the defense industry—leading on the one hand to a positive demand-shift for defense, and the completion of a major goal in the “Global War

on Terror” on the other—but especially because of their (plausibly) unexpected natures.<sup>3</sup>

To estimate causal effects, this paper estimates Bayesian structural time series (BSTS) models (e.g., Brodersen et al., 2015) at the firm level to generate synthetic controls (e.g., Abadie, Diamond and Hainmueller, 2010, 2015) of counterfactual financial performance. Causal effects at the firm level are estimated through comparisons of observed financial returns with estimates of their post-event trajectories from the statistical model. Counterfactual time series are estimated by leveraging information on each firm’s pre-event financial behavior and their associated relationship to “unexposed” firm returns in the pre-event window. The Bayesian approach provides several statistical advantages in this analysis. It allows both for meaningful posterior inference at the firm level and the relaxation of controversial identifying assumptions commonly invoked in event-study designs—such as the parallel trends assumption in a “difference-in-differences” design (Abadie, 2005), or that all causal effects be estimated with identical groups of covariates or control series, as is common in “market model” approaches (Kothari and Warner, 2008). Moreover, in the context of high-volume, daily financial data, this paper argues that the proposed model outperforms the traditional synthetic control estimator proposed in Abadie, Diamond and Hainmueller (2010, 2015), and even permits for traditional synthetic control estimates to be used in tandem with the Bayesian structural model. Evidence for this claim arises from a comparison of in-sample model fit of the competing methodologies, in addition to a series of simulation studies that compare the performance of the estimators under varied conditions.

In addition to providing novel firm-level causal estimates, the aggregation of individual event studies reveal heterogenous treatment effects and the possibility for varied political incentives within the defense sector. While these events are shown to shift the valuations (i.e., market capitalizations) of firms—with a typical company’s valuation

---

<sup>3</sup>At the time of Bin Laden’s killing, an *Intrade* betting market estimated the probability the terrorist leader ever being caught at lower than a 10% probability. The events likely vary in terms of their degree of “surprise,” however—the implications of which are discussed at length in Section 1.3.

shifting by as much as 5-10% as a result of individual events—not all defense firms appear to be influenced “in the same direction.” Results of the paper have broad implications for research on special interest groups and their relation to the broader political economy. Broadening our understanding of the challenges uniquely faced in the defense sector, estimates of firm-level effects make headway on the degree to which private contractors benefit or are hindered from large-scale, international events. Progress in this area may help us to understand the so-called “peculiar” dynamics (e.g., Drutman, 2015) of firm behaviors in the defense sector and the degree to which individual stakeholders, given the products they produce, are influenced by the international security environment.

The remainder of this paper proceeds as follows. Section 1.2 introduces the theoretical challenge of drawing valid causal inferences about the role of government expenditures and the international security environment on firm valuations. Section 1.3 discusses the chief empirical strategy used throughout the paper, the BSTS, and discusses competing approaches. Section 1.4 presents the main results of the analysis and discusses the estimates of individual effects across the events, cross-sectional heterogeneity, in addition to the relative performance of the competing approaches. To validate the performance of the BSTS model, a set of simulation studies are presented in Section 1.5. Section 1.6 discusses the overall implications of the study and paths for future research.

## **1.2 The Disconnect Between Firm Valuations, Defense Procurement, and War**

Since fiscal year FY2000, the U.S. House of Representatives appropriated nearly 5 Trillion USD to the DoD—the largest sum of money allotted to any U.S. federal department over that interval.<sup>4</sup> To put this total in perspective, it is roughly twice the size of all foreign

---

<sup>4</sup>This figure includes totals through conventional budget appropriations, emergency, and supplemental appropriations. Source: Federal Procurement Data Center. Not all of these contracts were awarded as

aid expenditures made by all governments around the world in the second half of the twentieth century (Easterly, 2003), and approximately eight-to-ten times the nominal cost of the Federal Reserve’s policy for “quantitative easing” in the aftermath of the 2008 Great Recession (Boesler and McCormick, 2016). In present dollars, the DoD averages over \$400 billion dollars in obligations on an annual basis, or more than a billion in contract expenditures per day. Since FY2000, about 25 million DoD contract actions (i.e., purchase orders, new awards, emendations to awards, or status changes) have been negotiated between the DoD and individual firms; more than 340,000 different companies won at least one DoD contract; more than 50,000 unique companies won awards valued at more than 1 million dollars or more; hundreds of companies with foreign headquarters won awards; more than 2,000 publicly-traded companies received contracts; and approximately 40 percent of all contract dollars were awarded in non-competitive bids.<sup>5</sup>

Despite the considerable data we have at our disposal, there is no straightforward way to simply observe how shifts in the international security environment may (instantaneously) help or hinder individual firms. Consider the visual presented in Figure 1.1, which plots financial data for a subset of prominent defense contractors across our three events of interest. The horizontal axis in each subplot maps trading days, and the vertical axis corresponds to a company’s normalized return—in this case, a firm’s average market capitalization on a given day divided by its market cap on the last day before the event. Gray lines at the center of each subplot mark the growth (in percentage point terms) between the value of the firm prior to the event and just following the event. The figure highlights how individual firms vary both in their post-event financial trajectories but

---

“wartime expenses.” A 2014 Congressional Budget Office (CBO) report estimates that at least 1.208 USD trillion was spent directly on the wars in Iraq and Afghanistan, although these totals are unlikely to capture all costs related to those conflicts (e.g., Belasco, 2014).

<sup>5</sup>These figures are author’s own tabulations and do not count dollars spent at the U.S. Department of State, or non-DoD agencies. If a contract is “non-competitive,” this means it is impossible for the government to solicit bids for a product in a blind or competitive manner, due to the fact that the nature of the good may be unique to individual firms.



also in their pre-event behaviors. Simply because a company's valuation rises or falls after an event does not reveal the effect of an event on an individual asset, however. To understand how an event impacts a firm requires a belief about how firm financials would have gone had the event not taken place. We detail our strategy for estimating firm-level counterfactuals in the next section.

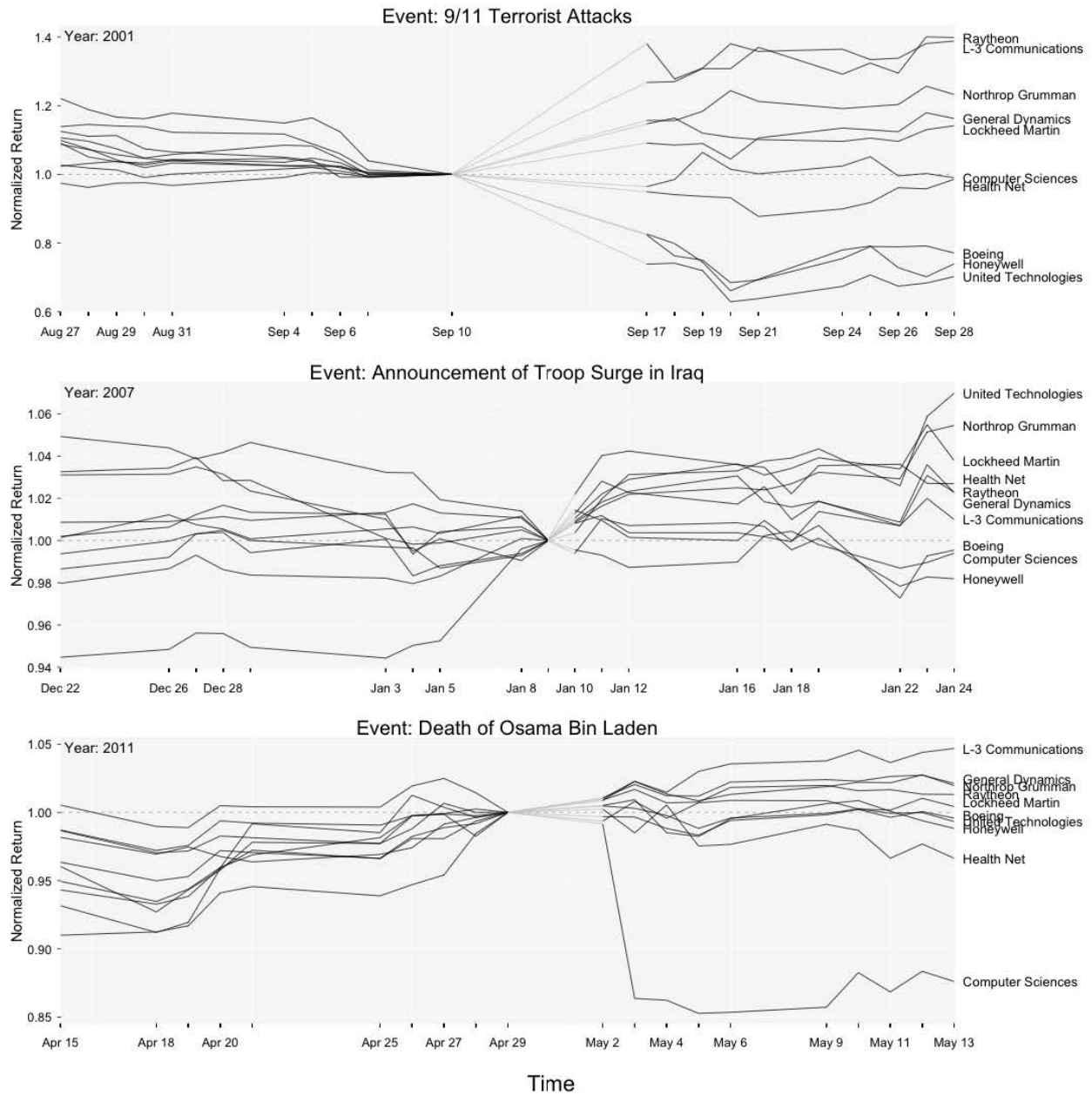
## 1.3 Empirical Strategy

### 1.3.1 Notation and Framework

The notation here is implied at the level of the individual event study. With minor emendations, the setup resembles that in Abadie, Diamond and Hainmueller (2010). In a given event study sample, there are  $N = n_0 + n_1$  total firms, where  $n_0 = N - n_1$  is the total number of "control" firms and  $n_1$  is the total number of hypothetically "exposed" (or treated) firms. A firm is considered exposed if had been awarded any number of DoD prime-contracting dollars in the two years prior to the event of interest. Time is discrete with  $T$  total periods and  $T_0$  pre-intervention (i.e., pre-event) periods, with  $1 \leq T_0 < T$ . Let  $Y_{it}^{obs}$  be exposed firm  $i$ 's observed market capitalization (i.e., its total public shares outstanding multiplied by share price) in time period  $t$ , for  $i = 1, \dots, N$ , and  $t = 1, \dots, T$ . The values  $Y_{it}^1$  and  $Y_{it}^0$  denote what firm  $i$ 's outcomes would have been under exposure and no exposure, respectively. Note that for firms that are truly exposed, the value  $Y_{it}^{obs} = Y_{it}^1$  while  $Y_{it}^0$  is unobserved; for control firms,  $Y_{it}^{obs} = Y_{it}^0$  while  $Y_{it}^1$  is unobserved. To lighten notation we may denote  $Y_{it}^{obs} = Y_{it}$  when appropriate. Unless otherwise noted, time periods are trading days not calendar days.

Our statistical aim is to estimate how exposed company valuations performed relative to their expectations following an event in time period  $T_0 + 1$  (e.g., counterfactual market

**Figure 1.1: Differing Trajectories of Observed Firm-Level Financial Returns After Major Events**



*Notes:* The dark, solid lines in each subplot represent normalized returns (i.e., a ticker’s closing price in a given period divided by its price on the final day before the event) for a given company ticker. The lighter lines at the center of each subplot indicate the timing of the event; lines become dark again on the first trading day after the incidence of the event. Recall that after the attacks on 9/11 trading markets were closed for one week.

performances at the firm level). In an ideal case, we could observe the quantity

$$Y_{it}^1 - Y_{it}^0, \quad (1.1)$$

which represents the period-level or “point-wise” difference in potential outcomes between exposure and no exposure for unit  $i$  in time  $t$ . Of course, unit-level comparisons of  $Y_{it}^1$  and  $Y_{it}^0$  are not possible due to the “fundamental problem of causal inference” (Holland, 1986)—namely, for any given unit in the sample, a unit is either exposed or not exposed to the treatment, which insures that at least either  $Y_{it}^1$  or  $Y_{it}^0$  will be unobserved in a given time period. Let us assume the observed outcome for unit  $i$  at time  $t$  can be written as:

$$Y_{it}^1 = \phi_{it} \cdot D_{it} + Y_{it}^0 \quad (1.2)$$

The treatment indicator  $D_{it}$  is defined as  $D_{it} = 1$  if  $i = 1$  and  $t > T_0$ , and 0 otherwise. As above, the treatment effect  $\phi_{it}$  can be defined as  $\phi_{it} = Y_{it}^1 - Y_{it}^0$ . The observed outcome for unit  $i$  and time  $t$  can be written as  $Y_{it} = \phi_{it} \cdot D_{it} + Y_{it}^0$ . The goal is to estimate  $\phi_{it}$  for  $t > T_0$ . We know that  $\phi_{it} = Y_{it}^1 - Y_{it}^0 = Y_{it} - Y_{it}^0$ .  $Y_{it}$  for  $t > T_0$  is observed, but in order to calculate  $\phi_{it}$  we need to estimate  $Y_{it}^0$ . The next subsection will describe our strategy for estimating this counterfactual quantity.

### 1.3.2 Statistical Model: Bayesian Structural Time Series with Synthetic Controls

The inferential challenge faced in this setting is how to reasonably estimate counterfactual firm financial performances. This section describes an approach for estimating the effect of an economic event on the value of individual firms. As the previous subsection highlights, the naïve comparison of company financials before and after an event is unlikely to identify the causal effect of the event on a company’s valuation because the post-treatment, counterfactual time series is unobserved at the unit level. Aside from the

trivial case in which the underlying time series is known to be stationary, a difference in means from before and after the event will be in expectation a biased estimate of the causal effect of an event on a firm's valuation. In practice, financial time series rarely satisfy stationarity assumptions and exhibit a great deal of serial correlation (Lewellen, 2002; Lo and MacKinlay, 1990).

Given that our causal quantities of interest are comparisons of observed returns against expected returns at the firm level (i.e., the behavior of an asset in absence of the treatment or "event"), a model is required to make reasonable estimates of these counterfactual time series. Popularized in statistical finance research, such methods are often referred to as "event study" designs. Several modeling choices are prominent in this literature, which range from traditional difference-in-difference approaches, comparative case study approaches, to "market model" approaches common in empirical finance. These techniques represent a suite of possibilities available to researchers interested in event study methods. See Campbell, Lo and MacKinlay (1997) and MacKinlay (1997) for more detailed discussions on the relative appeals and shortcomings of various techniques.<sup>6</sup>

To understand differential firm response to major events in U.S. foreign policy, this paper adopts the procedure presented in Scott and Varian (2014) and Brodersen et al. (2015) to fit a set of structural "state space" equations to estimate (post-event) counterfactual time series.<sup>7</sup> A non-technical introduction to this approach is discussed in Varian (2014). The appeal of this estimation approach is that we can estimate firm-level counterfactual series analogous to the influential "synthetic control" approach presented in Abadie and Gardeazabal (2003) and Abadie, Diamond and Hainmueller (2010, 2015), perform model selection on a set of potential covariates (i.e., control financial series), all the while allowing for valid posterior inference under a range of hypothetical data

---

<sup>6</sup>Prominent examples of event studies in pertinent to political research include Claessens, Feijen and Laeven (2008), Fisman (2001), Gaikwad (2013), Herron (2000), Jayachandran (2006), and Roberts (1990).

<sup>7</sup>An alternative approach to Bayesian structural estimation of counterfactual time series can be found in Street et al. (2015).

generation processes. This section presents the primary advantages and assumptions of the proposed modeling strategy.

At the firm level in the pre-event window, we estimate equations of generic structure:

$$\begin{aligned} Y_{it} &= Z_{it}^T \alpha_{it} + \varepsilon_{it}, & \varepsilon_{it} &\sim N(0, \sigma_{it}^2), \\ \alpha_{i,t+1} &= K_{it} \alpha_{it} + U_{it} \eta_{it}, & \eta_{it} &\sim N(0, Q_{it}). \end{aligned} \quad (1.3)$$

These two equations together fully describe the dynamics of a “state space” model, where the first equation is the observation (or measurement) equation and the second equation is the state (or transition) equation. Analogous notation can be found in Brodersen et al. (2015) and Durbin and Koopman (2002). The terms  $Q_{it}$ ,  $K_{it}$ ,  $U_{it}$ , and  $Z_{it}$  fully capture the state components and can be adapted to a wide range of dynamic settings. In fact, it can be shown that any (vector) auto-regressive moving average, integrated, or stationary process can be characterized in terms of observation and state equations of Equation 1.3 (see, e.g., Scott and Varian, 2014). This can be seen by considering the relationship between the outcome  $Y_{it}$  and the state vector  $\alpha_{it}$ . Note that if the relationship between  $Y_{it}$  and  $\alpha_{it}$  is fixed over time, this is equivalent to the case in which  $\alpha_{it} = 1$  and  $Z_{it} = \beta_i^T \mathbf{X}_{it}$ , or  $Y_{it} = \beta_i^T \mathbf{X}_{it} + \varepsilon_{it}$ . If the relationship between covariates and the outcome varies over time, however, the structure of the relationship can be specified with time-varying coefficients.<sup>8</sup>

In our main analysis, for each exposed firm and for each individual event studied, we estimate the following system of equations:

$$\begin{aligned} Y_{it} &= \mu_{it} + \beta_i^T \mathbf{X}_{it} + \varepsilon_{it} \\ \mu_{it} &= \mu_{i,t-1} + \delta_{i,t-1} + u_{it} \\ \delta_{it} &= \delta_{i,t-1} + v_{it} \end{aligned} \quad (1.4)$$

---

<sup>8</sup>Additional details on the mechanical differences between the static and time-varying coefficient models can be found in Brodersen et al. (2015). A possible concern with estimating time-varying coefficients, as the authors note, is that may be more prone to “overfitting” in the periods just prior to the event. For this reason we assume static coefficients for our main analysis.

where  $\beta_i^T \mathbf{X}_{it} = \sum_p \beta_i^p x_{it}^p$ ,  $p$  is the number of potential covariates to include in the model, and the error terms  $(\epsilon_{it}, u_{it}, v_{it})$  are assumed to be drawn from independent, mean-zero normal distributions. The parameter  $\mu_{it}$  captures a local linear trend in the outcome independent of the marginal influence of the set of potential controls, which may be influenced by correlations in errors across time periods. As the notation emphasizes, the model embeds fixed, autoregressive, and moving average components, as its objective is to be flexible and allow for the possibility of local-level trends and correlated errors across time periods. The control matrix,  $\mathbf{X}_{it}$ , includes (in theory) any time series that is predictive of the outcome series in the pre-event study window, but whose outcome is uninfluenced (relative to exposed firms) by the incidence of the event. In the context of financial returns, this assumes that control firms are not impacted by the timing of the event by means other than through period-specific common shocks.<sup>9</sup>

Ideal candidates for generation of the counterfactual series are any time series that are meaningfully predictive of the observed outcome in the pre-event window but are not exposed to the treatment; this includes, in theory, how well the outcome series predicts itself in the pre-event window, as compared to its reflexive relation in the post-event

---

<sup>9</sup>To illustrate this point further, consider the following data generation processes: exposed returns are generated by the series

$$Y_{it}^1 = \mu_{it} + \beta_i^T \mathbf{X}_{it} + \gamma_t + \mathbf{1}\{t > T_0\} \cdot \phi_{it} + \varepsilon_{it}^1,$$

while unexposed returns are generated by the series

$$Y_{it}^0 = \mu_{it} + \beta_i^T \mathbf{X}_{it} + \gamma_t + \varepsilon_{it}^0,$$

with  $\varepsilon_{it}^0, \varepsilon_{it}^1 \sim N(0, \sigma^2)$ . In this case  $\gamma_t$  is a period-specific common shock to both exposed and counterfactual series, and  $\mathbf{1}\{t > T_0\} \cdot \phi$  is an indicator function for if  $t > T_0$  (i.e., the current time period is after the event). In the pre-event period (i.e.,  $t \leq T_0$ ), the expected difference in outcomes between the exposed and unexposed series is  $\mathbb{E}(Y_{it}^1 - Y_{it}^0 | t \leq T_0) = 0$ . After the event, the expected difference in outcomes is  $\mathbb{E}(Y_{it}^1 - Y_{it}^0 | t > T_0) = \phi_{it}$ . Note that our unit-level, period-specific causal effects are given by  $\mathbb{E}(Y_{it}^1 - Y_{it}^0 | t > T_0) - \mathbb{E}(Y_{it}^1 - Y_{it}^0 | t \leq T_0) = \phi$ , where we see the common shock  $\gamma_t$  being differenced out. Here we see the existence of period-specific common shocks, which may be of concern in the context of major financial events, does not bias our estimates of the direct effect of exposure in this context—a fact that is true even if  $\gamma_t$  undergoes a secular shift after the event, so long as such a shift is common between the exposed unit and the counterfactual series.

window, if the post-event window were only a single period.<sup>10</sup> So long as control units do not systematically receive the direct effect of the exposure to the event and the exposure effect is independent of period-specific fixed effects common to exposed and unexposed units, the performance of post-event control series may be used draw inferences about the likely value of the counterfactual series (i.e., the value of the asset had the event not taken place). A similar discussion on the traits that make for ideal control variables can be found in Abadie, Diamond and Hainmueller (2010, 2015).

The familiar stable unit treatment value assumption, or SUTVA, (e.g., Rosenbaum and Rubin, 1983) is required to make unbiased causal inferences in this setting: namely, that a firm’s exposure to an event does not alter the probability its synthetic control will become differentially exposed to the event. While ultimately untestable, this assumption seems reasonable in our applied context given both the “surprise” nature of the analyzed events (e.g., both exposed and unexposed firms are unlikely to have restructured their product portfolios in anticipation of the event, or expended any effort to directly cause the existence of the event) and that the set of hypothetical control series are drawn from the subset of unexposed S&P500 firms (i.e., large-cap companies). In the event the non-interference assumptions fails, there is reason to believe the model will tend to underestimate the true firm-level treatment effects (Meyer, 1995; Rosenbaum, 2007).<sup>11</sup>

---

<sup>10</sup>By a similar logic, one could include firm-level, DoD contracting dollars as a control variable; however, if the time series has an immediate effect on short-term government purchases or new contract awards, the inclusion of such series could attenuate estimates of effects in the post-event window. Hence, in this analysis, the matrix of candidate controls is drawn from the set of S&P500 firms that received no DoD contract obligations in the prior two fiscal years.

<sup>11</sup>To illustrate a case in which a failure of SUTVA may bias our estimates, consider  $Y_{it}^0 = \mu_{it} + \beta_i^T \mathbf{X}_{it} + \gamma_t + \mathbf{1}\{t > T_0\} \cdot \rho \cdot \phi + \varepsilon_{it}^0$ , where  $\rho \in [0, 1]$ , but exposed series  $Y_{it}^1$  is generated as in the previous example. In this context the parameter  $\rho$  may be thought of as the degree to which the presumed-unexposed unit is actually exposed to the treatment, or the degree of dependence in potential outcomes. In this this context,  $\mathbb{E}(Y_{it}^1 - Y_{it}^0 | t > T_0) - \mathbb{E}(Y_{it}^1 - Y_{it}^0 | t \leq T_0) = \phi - \rho \cdot \phi = (1 - \rho) \cdot \phi$ . In this manner, if control series are actually exposed (even in part) to the incidence of the event, estimated treatment effects will be biased toward zero by the factor  $\rho$ . A rich discussion on the implications of SUTVA in settings such as our own can be found in Rosenbaum (2007).

## Inference

Posterior inference is performed through Markov Chain Monte Carlo (MCMC) utilizing the Gibbs algorithm discussed in Brodersen et al. (2015). This algorithm, with minor emendations, extends the popular approach first provided in Durbin and Koopman (2002). To estimate firm-level effects, we utilize the predictive relations between firm-level financial series and estimated model parameters in the pre-event window to derive the posterior distribution of unit-level effects in the post-event window. Let  $\theta$  capture the set of model parameters in Equation C.5. Values of the local linear trend and coefficients in Equation C.5 are simulated given the observed sequence  $\{Y_{it}\}_{t=1}^{T_0}$ . Next, the posterior predictive distribution of  $p(\{\widehat{Y}_{it}\}_{t=T_0+1}^T | \{\widehat{Y}_{it} | \theta\}_{t=1}^{T_0})$  is derived taking the posterior estimates derived in the preceding step. Across all sample periods  $t = 1, \dots, T$ , the posterior predictive distribution of point-wise causal  $Y_{it} - \widehat{Y}_{it}$  effects is equivalently attained. See Durbin and Koopman (2002) and Brodersen et al. (2015) for additional detail on the Gibbs algorithm utilized herein.

Estimates of the point-wise (or period-specific) causal effect from a single draw from the Markov Chain  $d$  are given by:

$$\widehat{\phi}_{it}^{\{d\}} = Y_{it} - \widehat{Y}_{it}^{\{d\}}. \quad (1.5)$$

where an estimate of the expected value of the effect,  $\phi_{it}$ , is given by the average difference across draws. A total of 2500 draws from the Markov Chain are taken to derive each firm-level estimate. Averaging over the post-event window, an estimate of the firm-level, running-average causal effect is given by

$$\bar{\phi}_i = \frac{1}{T - T_0} \sum_{t=T_0+1}^T \phi_{it}. \quad (1.6)$$

When firm-level outcomes are scaled in terms of abnormal returns, estimates of the



period-specific, buy-and-hold abnormal return (AR) effect are given by:

$$AR_{it'} = \prod_{t=T_0+1}^{t'} \left[ 1 + \frac{Y_{it} - Y_{i,t-1}}{Y_{i,t-1}} \right] - \prod_{t=1}^{t'} \left[ 1 + \frac{\widehat{Y}_{it} - \widehat{Y}_{i,t-1}}{\widehat{Y}_{i,t-1}} \right], \quad (1.7)$$

with its associated running-average effect given by:

$$\overline{AR}_i = \frac{1}{T - T_0} \sum_{t=T_0+1}^T \left( \prod_{t=T_0+1}^T \left[ 1 + \frac{Y_{it} - Y_{i,t-1}}{Y_{it}} \right] - \prod_{t=T_0+1}^T \left[ 1 + \frac{\widehat{Y}_{it} - \widehat{Y}_{i,t-1}}{\widehat{Y}_{it}} \right] \right). \quad (1.8)$$

It should be emphasized that  $\overline{AR}_i$  has an important strategic interpretation for a hypothetical shareholder: it is an estimate of the expected difference (in percentage point terms) in a financial return that is attributable to the incidence of the event over the post-event window (Barber and Lyon, 1997; Fama, 1998; Kothari and Warner, 2008). Just as with the period-specific effects, posterior inference on  $\overline{AR}_i$  is obtained by considering the distribution of effects across draws of the MCMC algorithm.

### 1.3.3 Selection of Control Variables

As described in the previous subsection, our estimation procedure allows for our control series to vary at the firm level. A common procedure in event-study methods is to calculate abnormal firm returns (after an event) with respect to an aggregate market index, such as the S&P500 or the Dow Jones Industrial Average. While this procedure is attractive due to its relative simplicity, it may lead to biased inferences about the expected behavior of an asset (absent the event) if an exposed firm's returns are used to calculate the benchmark index. This concern appears to be valid in our applied context, as roughly half of all S&P500 firms had been awarded at least one DoD contract across our event studies of interest. As of June 2016, for example, roughly 3% of the mass of the S&P500 was determined by the financial performance of Lockheed Martin. To perform an event study on Lockheed Martin using the S&P500 as a control variable would necessarily distort estimates derived from an event study, as Lockheed Martin's financial returns be

explained by a function of itself.

Given a set of potentially many control covariates, we embed, as in Ishwaran and Rao (2005) and Scott and Varian (2014), a “feature selection” step to the estimation of Equation C.5 to attempt to account for which control series, if any, to include in each firm-level model. In a single draw of the Markov Chain, the “inclusion” of an individual control series is given by its posterior probability of having a non-zero coefficient estimate in the observation equation. The popular “spike-and-slab” prior distribution (Ishwaran and Rao, 2005; Mitchell and Beauchamp, 1988) is utilized for weighting hypothetical regression estimates in the final model. The “spike” refers to the case that coefficient estimates have prior weights towards zero, while “slab” refers to the degree of informativeness of the prior, ranging from greatly spiked (i.e., weighted toward zero) to flat (i.e., a wide-tailed Gaussian prior). A rich discussion of the structure and merits of the spike-and-slab prior for state space models is provided by Durbin and Koopman (2002).<sup>12</sup> In this analysis we similarly assume an indifference prior—namely, that each variable’s inclusion in the final model can be modeled as independent Bernoulli draws with an priori probability of 50% (George and McCulloch, 1993; Ishwaran and Rao, 2005), and the estimation algorithm follows Brodersen et al. (2015). Additional information on the results of this procedure can be found in Appendix A.4.

In contrast with the algorithm presented in Abadie, Diamond and Hainmueller (2010, 2015)—where counterfactual series are generated from convex combinations of unexposed outcome series—the model selection technique utilized here has several distinctive features. In the proposed Bayesian framework, exposed outcome series are not required to exist within the convex hull of the set of potential control series. In the event no controls are found to be meaningfully predictive of the observed series in the

---

<sup>12</sup>Other prominent feature selection techniques (e.g., lasso, regularized regression) could be used at this stage, and there are close relationships between Bayesian feature selection and the lasso (e.g., Park and Casella, 2008). The implicit gain of the embedded procedure is that it appropriately adjusts model uncertainty for the feature selection process.

pre-event window, no controls need to be included in the model whatsoever to generate causal estimates. Unlike the traditional synthetic control estimator, the estimation of unit-level causal effects is possible even if no control series suitably predict the exposed series in the pre-event window. Another way to state this is that counterfactual series can be generated even in absence of unexposed control firms. This flexibility comes, of course, at the cost of assuming a parametric form to the data generation process; but this seems like a worthwhile tradeoff in the context of financial time series, as one might have prior institutional knowledge that can be utilized to construct informative priors, or wish to estimate models explicitly across a particular lag structure—all choices that may improve the efficiency of the model.<sup>13</sup>

### **Baseline Controls: Unexposed Market Indices and Principal Components**

Rather than use a composite index like the S&P500 to benchmark individual returns, we construct event-study specific market indices from the subset of unexposed S&P500 constituent firms that were publicly traded across the full event study period. In a given event study, this leaves roughly unexposed 250 firms as potential control variables. While in principle one could estimate the model in Equation 1.6 using the full set unexposed S&P500 firms—or even utilize all unexposed firms, regardless of their being a part of a market index—as potential controls, we use a far more parsimonious model throughout our analysis. We do this for two primary reasons: first, reducing the number of potential controls greatly reduces computation time, and second, almost all of the variation of these time series can be explained by a much smaller representation of the data. Dimension reduction, in other words, appears to come at little inferential cost. As such, we utilize only 11 covariates in the baseline version of the model. In the space below we will

---

<sup>13</sup>Moreover, in the event no hypothetical control series meaningfully predicts the exposed series in the pre-event window, individual-level inferences are still possible by utilizing estimates from the local-linear parameter.

describe their generation and rationale.<sup>14</sup>

Define the set of unexposed S&P500 constituent firms in time  $t$  as  $C$ . For each event study, we create a new market index

$$\text{Market}_t = \sum_{i \in C} Y_{it}, \quad (1.9)$$

which is simply the sum of the market capitalizations of each firm  $i \in C$  in time  $t$ . This series resembles the normal S&P500 index—which is also a market-capitalization weighted index—but effectively removes the influence of firms that had received DoD money in their prior two fiscal years. Across event studies, there is a strong correlation between the (pre-event) levels of  $\text{Market}_t$  and the S&P500 index (i.e.,  $\rho > 0.95$  in all cases). While at first blush the strong association between these two indices would suggest it is acceptable to use the S&P500 index instead of  $\text{Market}_t$ , we nevertheless prefer the custom market index as a control variable.<sup>15</sup>

In addition to our creation of  $\text{Market}_t$ , we perform a Principal Component Analysis (PCA) to reduce the dimensionality of our unexposed collection of financial series. The theory of PCA has a long history in statistical research and applied problems in finance.<sup>16</sup> The aim of PCA is to take a (potentially high-dimensional) collection of variables and transform it into lower-dimensional representation that retains much of the variation (or predictive power) of the original data, where each “principal component” is a linearly-uncorrelated dimension. For our purposes, each of these components may be thought of as a potential control variable to include in our model, subject to the variable selection procedure outlined before. Without loss of generality, assume the first  $|C| = c$  firms of

---

<sup>14</sup>Note that no covariates are actually required to estimate Equation 1.6. In such a case, the model could nevertheless capture time-varying behavior with the autoregressive component  $\mu_{it}$ .

<sup>15</sup>After all, discrepancies between the two measures are precisely attributable to individual defense firms, not random noise.

<sup>16</sup>An early example of multi-factor analysis in finance comes in Lessard (1974). See Campbell, Lo and MacKinlay (1997) and Lai and Xing (2010) for general discussions on the merits of PCA for multi-factor analysis in financial research.

the  $N$  total are firms in  $C$ . Next, define this collection of control returns,  $\mathbf{Y}$ , as follows:

$$\mathbf{Y} = \begin{pmatrix} \tilde{Y}_{1,1} & \tilde{Y}_{1,2} & \cdots & \tilde{Y}_{1,c} \\ \tilde{Y}_{2,1} & \tilde{Y}_{2,2} & \cdots & \tilde{Y}_{2,c} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{Y}_{T,1} & \tilde{Y}_{T,2} & \cdots & \tilde{Y}_{T,c} \end{pmatrix},$$

where  $\tilde{Y}_{it} = (Y_{it} - \bar{Y}_{it})/s_i$  and  $s_i = \sqrt{\frac{1}{T_0-1} \sum_{t=1}^{T_0} (Y_{it} - \bar{Y}_{it})^2}$ . We perform PCA via a singular value decomposition as in  $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^T$ . The singular values of  $\mathbf{Y}$  are contained in the rectangular-diagonal matrix  $\mathbf{\Sigma}$ , while  $\mathbf{U}$  and  $\mathbf{W}$  are matrices capture the singular vectors and right singular vectors, respectively. All told, we retain only the first 10 principal components for use as potential control variables. Across event studies, the first 10 components account for well-above 90% of the observed variation in  $\mathbf{Y}$ . Call these variables  $PCA_1, PCA_1, \dots, PCA_{10}$ . In this specification, the set of potential variables  $\mathbf{X}_{it}$  includes the PCA variables and  $Market_t$ .

### **Alternate Specification: Incorporating Traditional Synthetic Controls**

Abadie, Diamond and Hainmueller (2010, 2015) propose a systematic, data-driven method for estimating a “synthetic control” counterfactual series in settings such as ours, which is achieved by, first, taking a weighted average of all possible control units, and second, building a synthetic unit that best matches characteristics of the pre-treatment treated-unit over a range of time. The main argument for synthetic controls is that may create better matches along pre-treatment characteristics than would traditional controls, or even matched pairs. This is achieved by analyzing the set of (unexposed) pre-treatment outcome variables and optimizing a weighting algorithm for a specified period of time.

If proper assumptions hold, the synthetic control should accurately represent the unobserved potential outcome of the treated unit in the post-treatment period; it should also control (in theory) for unobserved time-varying covariates. This depends, however,

on how well the synthetic control matches the treated unit. The authors propose the mean square prediction error (MPSE) in the pre-event window to measure how far the synthetic outcome is from the actual outcome. Qualitatively, this requires comparing values of predictor (and outcome) variables of synthetic and treated units, which is possible because of the explicit weights given to control units and covariates. This entire process is conducted without looking at the post-treatment data. If pre-treatment outcomes nicely match, then any difference in treated-unit and synthetic control post-treatment outcome measures can be considered the treatment effect. If pre-treatment outcomes do not nicely match, however, then subsequent analysis cannot be performed.

The authors demonstrate that under general conditions the value of the synthetic control unit should be estimated by  $\text{Synth}_{it} = \sum_{c \in C} \widehat{w}_c \cdot Y_{ct}$ , for  $i \notin C$ , where the vector of weights  $\widehat{w}$  is given by the solution to  $\widehat{w} = \arg \min_{\{w\}} \left[ (Y_i - w \cdot Y_c)' \Omega (Y_i - w \cdot Y_c) \right]$ , where the weight matrix  $\Omega$  minimizes the pre-event prediction error,  $Y_i$  is exposed firm  $i$ 's financial series, and  $Y_c$  is the set of financial series for firms in  $C$ . Firm series  $Y_i$  and  $Y_c$  are normalized by dividing by the values of  $Y_{i,T_0}$  and  $Y_{c,T_0}$ , respectively. While there may be numerous solutions to this problem, we follow the authors' computational recommendations and require that weights be non-negative. In suit, our alternate specification of the BSTS model, the set of potential control variables  $\mathbf{X}_{it}$  includes the PCA variables,  $\text{Market}_t$ , and  $\text{Synth}_{it}$ .

### 1.3.4 Study Windows and External Validity

Event study designs require the researcher define which units are exposed to an event, a means to estimate a unit's counterfactual series, and the pre- and post-event windows (Campbell, Lo and MacKinlay, 1997; MacKinlay, 1997). In general, there is no "silver bullet" for determining how long the pre- and post-event windows should be. If the post-event window length is too long, causal effects may be biased by the possibility of intervening (unobserved) treatments. If study windows are too short, the researcher

sacrifices statistical power and the possibility that treatment effects may not be lasting, or that their occurrence might not be immediately realized in financial markets. For each firm-level event study in this analysis, we therefore define the 500 trading days prior to the event as the pre-event window and the 20 days after the event as the post-event window.<sup>17</sup> This specification aims to draw causal inferences in the short- to medium-term, and allow for a rich sample of pre-event data to efficiently train our statistical models.

The estimation of unbiased causal effects in event studies requires that events be unanticipated by the marketplace. If expectations over an event have already been partially “priced in” to a given stock, estimates of firm-level causal effects will be driven towards zero as a function of traders’ prior (subjective) beliefs about the occurrence of the event. If a researcher has appropriate measures of traders’ a priori beliefs regarding the probability of an event, however, biased causal estimates can be transformed to recover unbiased effects (e.g., Snowberg, Wolfers and Zitzewitz, 2011). In practice, the existence of such measures is quite rare and in the event they do exist may presuppose that an event is already somewhat anticipated; and in absence of such measures, care is required in selecting events for study.

Aware of these concerns, we focus on three events in our analysis: the terrorist attacks of 9/11 (event date: 2001-09-11), President Bush’s announcement of a future troop “surge” in Iraq (event date: 2007-01-10), and the death of Osama Bin Laden (event date: 2011-05-02). These events undoubtedly vary in terms of their political significance, how they may be expected to shift the demand for defense services, and perhaps even their degree of “surprise.” 9/11 is widely considered to be one of the biggest foreign policy surprises in the history of U.S. foreign policy. Formal plans for a “troop surge” in Iraq were announced by President Bush on a weekday evening, a full two weeks prior to a vote

---

<sup>17</sup>In the Appendix we perform analyses that utilize varying post-event window lengths. We find that our choice of  $T - T_0 = 20$  versus  $T - T_0 = 100$  does not substantively alter our main findings.

being held in Congress over that issue.<sup>18</sup> The death of Osama Bin Laden occurred late in the after-hours market on a Sunday evening, before markets re-opened Monday morning. The first two of these events were chosen as (plausibly) exogenous cases of “increases” in the demand for defense services. The death of Bin Laden, by contrast, was selected as a case that might presage an eventual drawdown of troops in Iraq and Afghanistan, or lead to a possible demand reduction in the Global War on Terror.

The ability to generalize about how firms respond to events such as these depends, of course, on how similar the contexts of these events are to outside cases. For example, even if another large-scale terrorist attack were to occur on American soil, it might be misleading to assume that financial markets would respond just as they did on 9/11. This is not only because firm portfolios and government purchasing patterns may change over time, but also because beliefs over the occurrence of an event may not be constant across cases, if earlier events change the market’s beliefs about the possibility of a future event (e.g., Eldor and Melnick, 2004). Our aim in this analysis is therefore to make qualitative judgments about the behavior of defense firms after such cases, rather than make strict claims about how individual firms will necessarily respond to events outside the sample.

### 1.3.5 Data Sources

Financial market data were obtained from Compustat USA and the CRSP systems, accessible online through the Wharton Research Data System. Firm-level procurement records were accessed from public records via the Federal Procurement Data System (FPDS), tabulating all prime awards at the Department of Defense. Between FY2000 and

---

<sup>18</sup>Prior to the announcement of the troop surge in Iraq, there were roughly 20,000 fewer U.S. troops in Iraq relative to the year before, and sectarian violence was reported to be at an all-time high (Kagan, 2010). Given that many of these wartime details were publicly observable at the time, it is possible that firms could have partially anticipated the possibility of a troop surge prior to President Bush’s announcement. By the end of the year 2006, the White House had repeatedly conferred with military officials about the possibility of a major shift in strategy in the Iraq War. Most of these deliberations were made in private, but speculation of the possibility of a surge was public just before the new year (e.g., Keane and Kagan, 2006).



FY2015, roughly 25,000,000 contract actions were recorded in this database, and over 340,000 unique firms were awarded at least one federal contract. The total value of these contract awards nears 5 trillion USD. In general, firm names and numeric identifiers are not constant across datasets, which impedes the straightforward merging of these records. Firms in the FPDS are indexed with proprietary DUNS numbers provided by Dun & Bradstreet; other standard company identifiers, such as the CIK provided by the SEC, are not present in the FPDS data. Appendix A.15 details an algorithm that was used to match firm names across databases. Utilizing a list firm names taken from the Compustat and CRSP databases, matches in FPDS were determined by searching for exact and “fuzzy” name (substring) matches via the Levenshtein distance, after accounting for differences in word order, capitalization, punctuation, and common abbreviations in company titles (e.g., “Boeing Company” versus “THE BOEING CO”). Candidate matches were then manually verified to create firm-level mappings across databases. More detailed descriptions of the data utilized in this analysis and how data were cleaned can be found in Appendix A.3.

## **1.4 Results: Firm Financial Responses to 9/11 Attacks, the “Troop Surge” in Iraq, and Death of Osama Bin Laden**

### **1.4.1 Firm-level Results**

Results of the analysis indicate that major defense contractors vary markedly in terms of how they respond to major events. Figure 1.2 presents results of how a prominent set of firms responded to the 9/11 terrorist attacks, in addition to their estimated point-wise causal effects after the event. The top panel (A) marks a set of firms that are estimated to have been positively impacted by the terrorist attack. The lower panel (B) marks a set of prominent firms that felt negative shocks from the event. The top row of panel (A) plots a given firm’s observed financial series (—) against its estimated counterfactual series

from the baseline BSTS model (---). The gray-shaded region in each subplot indicates the 95% posterior predictive interval around each period-level estimate. As the figure shows, prominent defense firms such as Raytheon, Lockheed Martin are estimated to have experienced measurable increases in their valuations following the terror attacks (gaining between 2 and 5 billion USD each), and the effects appeared to be lasting for many trading days following the incidence of the event. Foundation Health Systems—at the time, a major healthcare provider to the U.S. armed forces, which later changed its name to Health Net—also appears to have benefited measurably from the 9/11 terror attacks. This observation suggests that major shifts in defense demand may impact firms over a broad range of specializations, not simply those that directly aid the process of war-fighting. If an event increases the probability of conflict, and an increase in the probability of conflict increases demand for health services for soldiers, surprise events may impact firms that superficially appear unrelated to a war effort.

Simply because a firm sells a high volume of defense-related goods to a government, however, does not imply that a firm will necessarily benefit from a major shift in the security environment. Evidence for this claim comes from looking at the financial response of firms such as A.A.R. Corp, Boeing, and United Technologies, which were negatively impacted by the incidence of the terrorist event, appearing to lose billions of dollars relative to their counterfactual series (at least in the short term). Much like Lockheed Martin, this particular set of “losing” firms yields high levels of contract awards for the construction of high-performance aircraft and materials used for military aviation; relative to Lockheed Martin, however, these firms financially depend much more heavily on commercial sales through non-defense channels. Given that the terrorist attacks of 9/11 were known to negatively influence airline stocks in the short-term due to sudden shifts in the demand for travel (e.g., Drakos, 2004), analysis of the event reveals that an individual firm’s post-event welfare is likely a function of both shifts in the demand for national defense but also an event’s effect on other non-defense revenue streams. In

other words, if a sudden shift in the international security environment simultaneously aids one aspect of a firm's portfolio but hurts another, the net effect of the event will be determined by the full set of risk factors that shape a firm's livelihood and the relative influence of defense versus non-defense sector demand shifts. Firms that appear to be superficially similar in terms of levels of procurement awards and defense-product specializations—such as Boeing and Lockheed Martin—may have divergent responses to the same event precisely because of their degree of financial dependence on non-defense sales.

Similar visuals depicting firm reactions to the Iraq troop surge are presented in Figure 1.3. The announcement of the troop surge benefited firms like L-3 Communications (a firm that specializes in military communications systems, personnel training, and facility management), Oshkosh Corporation (a primary supplier of military vehicles and speciality trucks to the U.S. armed forces), and Health Net (a prominent medical provider for U.S. soldiers and veterans). Note that across these graphs there is a slight “lag” between the incidence of the event and the measured shifts in company valuations. The timing of this delay coincides with U.S. House vote on the Iraq troop surge two weeks later, which perhaps suggests markets are more likely to respond to the formal agreement over a policy shift than less formal speeches concerning such strategies. But just as with the 9/11 event study, not all major defense contractors appeared to benefit from the announcement of the troop surge. Firms like Boeing, Computer Sciences Corporation (a major provider of information technology services to the Defense Department), and ITT Industries (which builds specialized components for aerospace and transportation products) experienced financial returns that were squarely within expected levels as forecasted by the BSTS model.

Lastly, Figure 1.4 visualizes firm-level results following the death of Osama Bin Laden. Firms like Northrop Grumman (a major defense contractor specializing in aerospace weaponry and logistics support services) and Level-3 Communications (a major contractor

with the Defense Information Systems Agency in the DoD) appeared to positively benefit from the terrorist's death. Firms such as Boeing had causal effects that were positive but statistically insignificant by conventional standards. Firms negatively impacted by the death of Bin Laden include Health Net, Harris Corporation (a firm that specializes in wartime tactical equipment and communications systems), and the Oshkosh Corporation. Despite specializing in different defense products, these "harmed" firms all benefitted from the Iraq troop surge. This relationship is consistent with the claim that Bin Laden's death acted as a negative shock for several firms, especially those that benefit from increased numbers of boots on the ground.<sup>19</sup>

### 1.4.2 Cross-Sectional Heterogeneity

A common procedure in event study designs with many exposed firms is to aggregate estimates of individual effects to understand broader market-wide trends (Campbell, Lo and MacKinlay, 1997). The most common practice in this setting is to run regressions with firm-level effects as the outcome variable and firm-level characteristics as the explanatory variables. While this procedure is somewhat impressionistic—as such regressions, in absence of detailed production functions, may be sensitive to appropriate functional forms being specified—the act may nevertheless shine line on the factors most strongly correlated with firm-level causal effects.<sup>20</sup>

This subsection details several approaches to inspecting the cross-sectional heterogeneity of individual causal effects. The primary approach compiles firm-level procurement

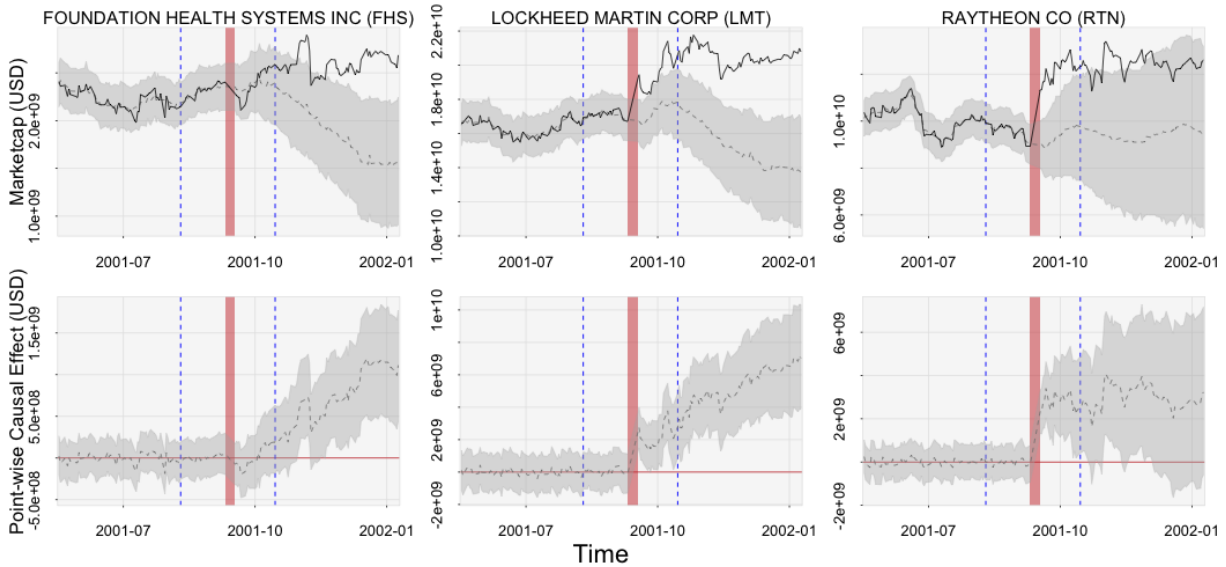
---

<sup>19</sup>The fact that some primary contractors benefited from Bin Laden's death while others appeared harmed is also consistent with a belief that reductions in military boots on may be associated with increases in private security workers (Peters, Schwartz and Kapp, 2015). Utilizing FPDS records on their own, however, there is no procurement code for active civilian personnel in battlefields, which makes it difficult to test this claim explicitly.

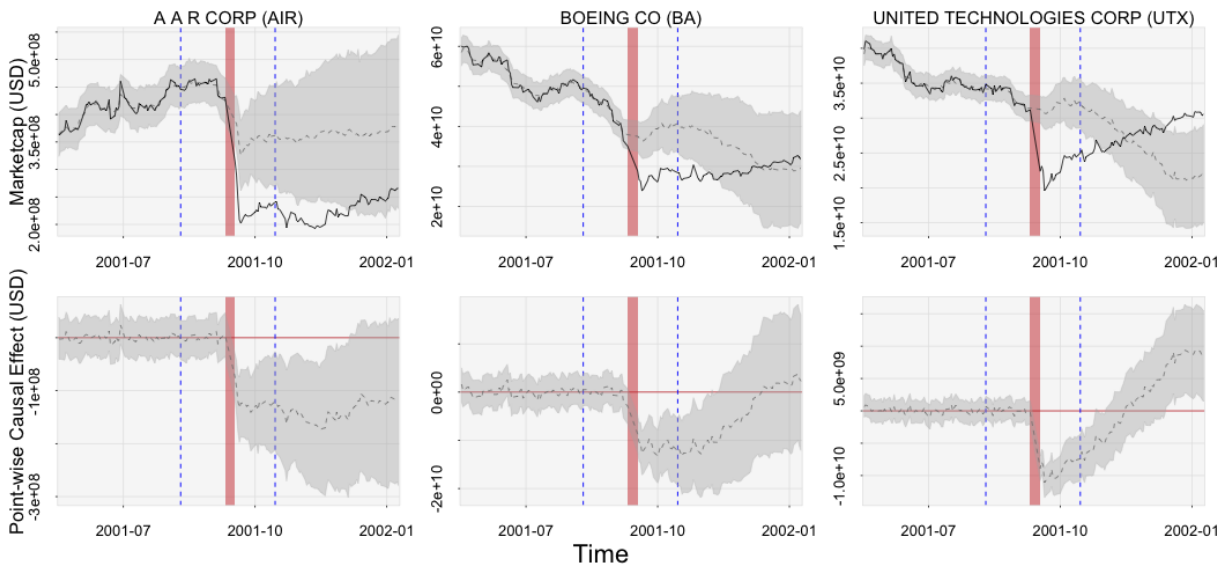
<sup>20</sup>Additional challenges with comparisons across studies come from the fact that different sets of firms are exposed across studies, firm portfolios change over time, and that events themselves may be qualitatively different. Different subsets of the defense sector may be influenced by some events and not others, in other words, and within-firm comparisons may be complicated by changing production portfolios over time.

**Figure 1.2: Firm-Level Responses to 9/11 Attacks**

*(a) Prominent Firms that Benefited from Event*



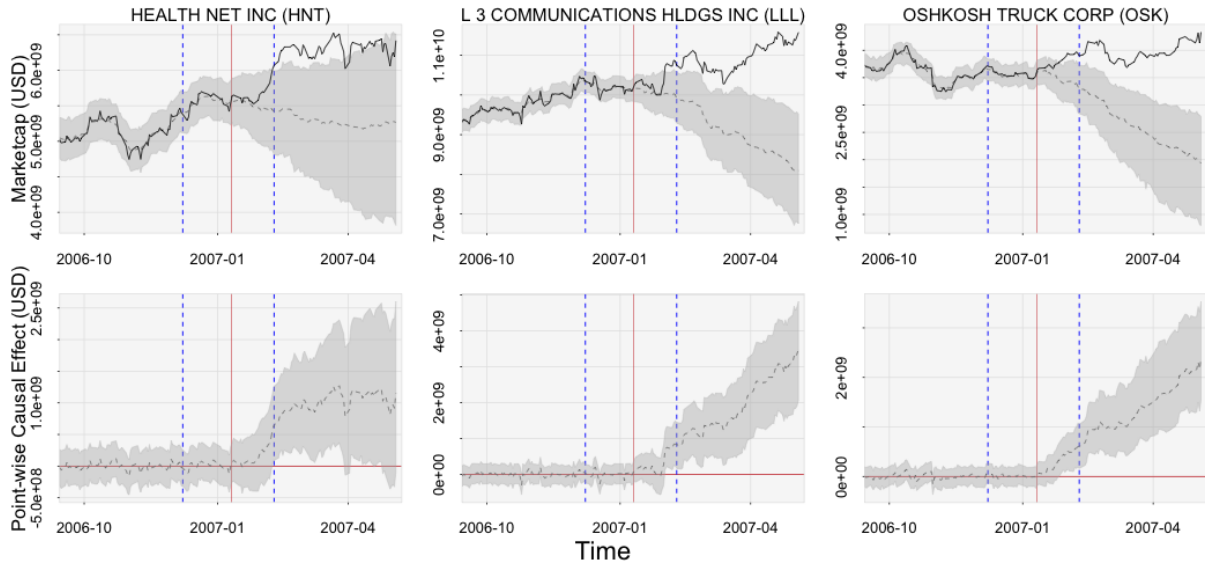
*(b) Prominent Firms that did not Benefit from Event*



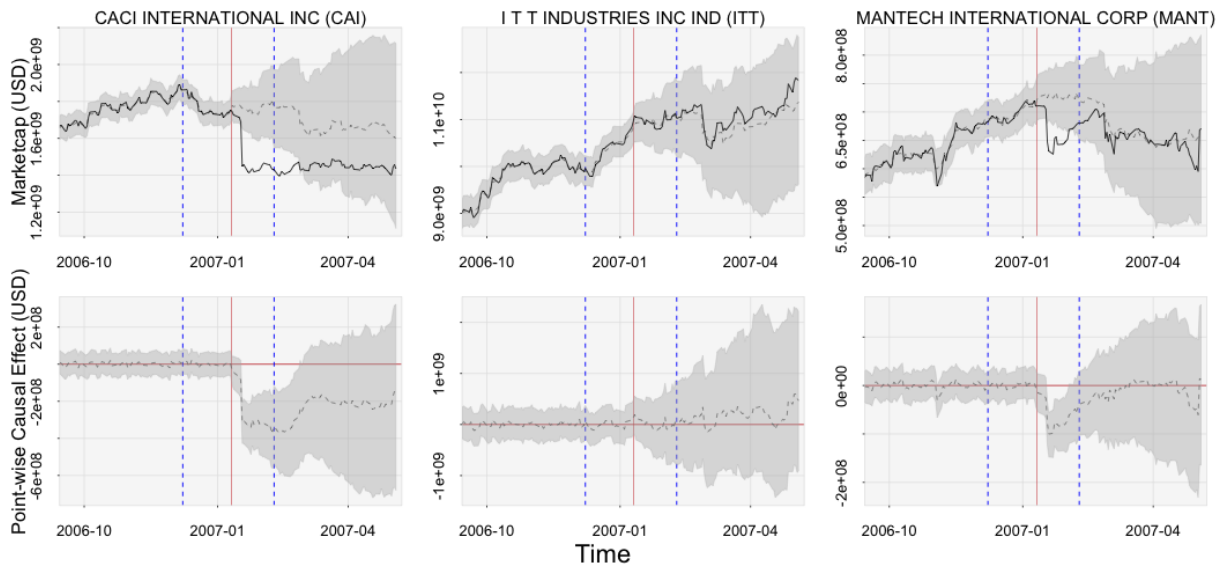
*Notes:* The top row in each panel compares the observed market cap for a given firm over time (—) against its estimated counterfactual (---). The shaded region represents a 95% prediction interval from the Bayesian structural time series. Vertical dotted-lines denote the twenty trading days before and after the event. Similarly, the lower row in each panel shows the observed market capitalization for a given firm minus its expected return (i.e.,  $Y_{it} - \hat{Y}_{it}$ ), with shaded regions denoting 95% posterior intervals. 500 trading days prior to the event are used to estimate models; for visual clarity, the range of each plot is limited to the 80 trading days before and after each event.

**Figure 1.3: Firm-Level Responses to News of Iraq Troop Surge**

*(a) Prominent Firms that Benefited from Event*



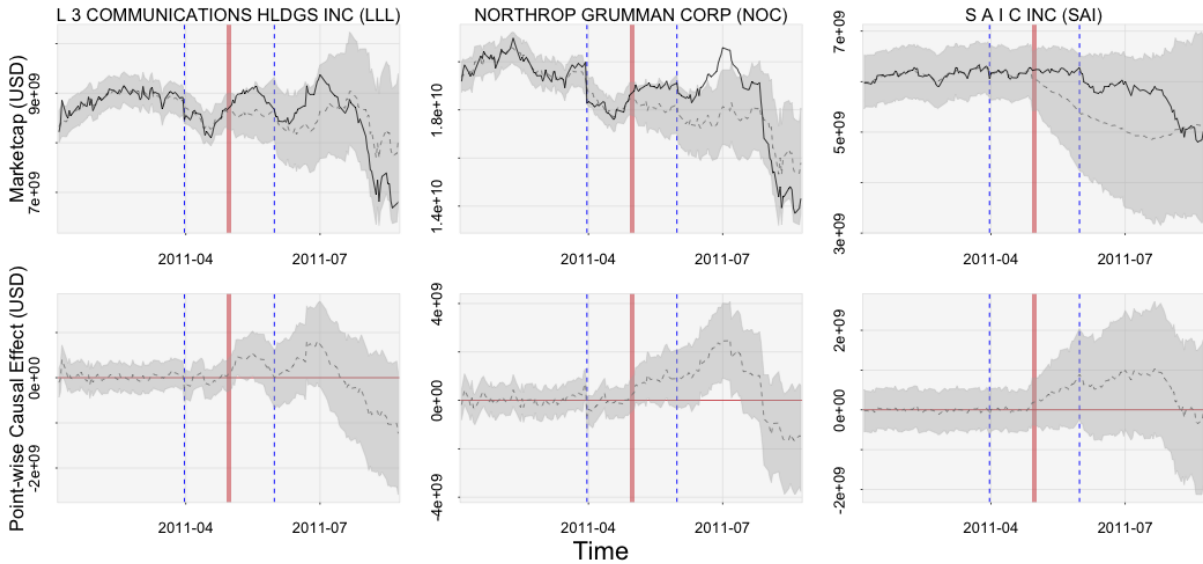
*(b) Prominent Firms that did not Benefit from Event*



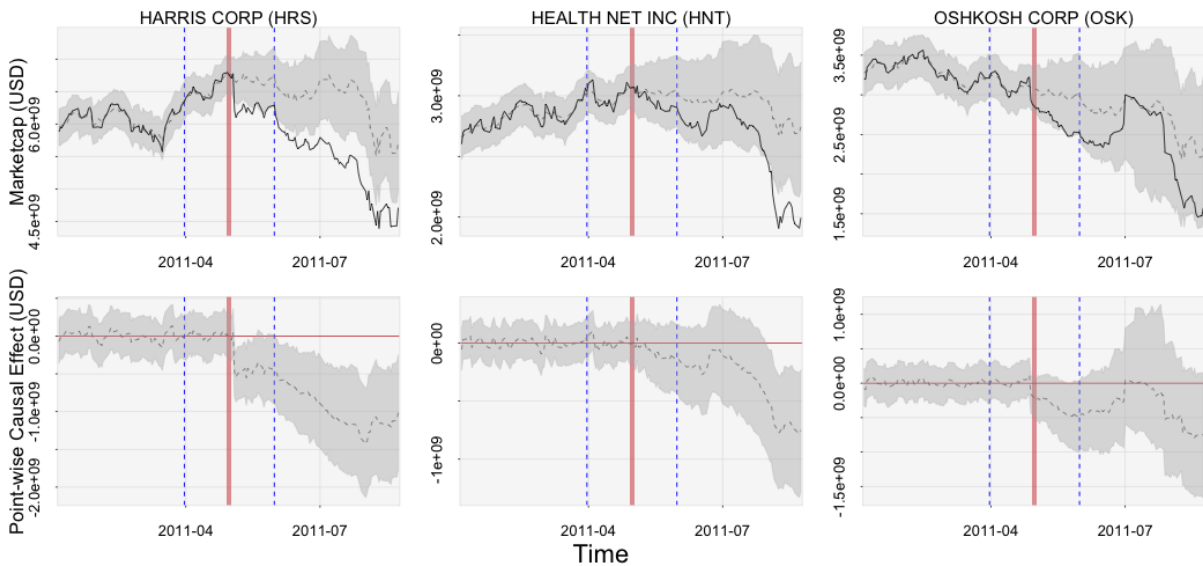
*Notes:* The top row in each panel compares the observed market cap for a given firm over time (—) against its estimated counterfactual (---). The shaded region represents a 95% prediction interval from the Bayesian structural time series. Vertical dotted-lines denote the twenty trading days before and after the event. Similarly, the lower row in each panel shows the observed market capitalization for a given firm minus its expected return (i.e.,  $Y_{it} - \hat{Y}_{it}$ ), with shaded regions denoting 95% posterior intervals. 500 trading days prior to the event are used to estimate models; for visual clarity, the range of each plot is limited to the 80 days before and after each event.

**Figure 1.4: Firm-Level Responses to Osama Bin Laden's Death**

*(a) Prominent Firms that Benefited from Event*



*(b) Prominent Firms that did not Benefit from Event*



*Notes:* The top row in each panel compares the observed market cap for a given firm over time (—) against its estimated counterfactual (---). The shaded region represents a 95% prediction interval from the Bayesian structural time series. Vertical dotted-lines denote the twenty trading days before and after the event. Similarly, the lower row in each panel shows the observed market capitalization for a given firm minus its expected return (i.e.,  $Y_{it} - \hat{Y}_{it}$ ), with shaded regions denoting 95% posterior intervals. 500 trading days prior to the event are used to estimate models; for visual clarity, the range of each plot is limited to the 80 trading days before and after each event.

records with publicly disclosed financial statements to scale a firm’s overall revenue dependence to the DoD.<sup>21</sup> To assess how firm-level causal estimates vary as a function of a firm’s financial dependence to the DoD, we construct firm-level measures utilizing data from the FPDS and company-level financial statements released to the SEC. Each financial quarter, publicly traded firms on U.S. markets are required to tabulate their total revenues from all their business dealings. As such, we merge quarterly firm revenues (taken from a firm’s 10-Q document) with disclosed contract obligations from the Department of Defense. Firm-level measures of financial dependence—which may be conceived of as a measure of exposure or “connection,” as in Fisman (2001)—are given by:

$$\nabla_i = \frac{\sum_{q \in Q} \sum_k d_{iq}^k}{\sum_{q \in Q} \text{Revenue}_{iq}}, \quad (1.10)$$

where  $d_{iq}^k$  are total contract dollars awarded to firm  $i$  for product  $k$  in quarter  $q$ , and  $Q$  is the set of 8 financial quarters preceding the event date. We only include firms with 8 quarters of reported SEC revenue data and with pre-event market capitalizations of over 100 Million USD. To assess how  $\nabla_i$  is associated in the aggregate with firm-level causal effects, we estimate quantile regressions as in Koenker and Hallock (2001) that solve the following:

$$\min_{\theta \in \mathbb{R}} \sum_{i=1}^n \pi_{\tau} \left( \overline{\text{AR}}_i - \mu(\nabla_i, \theta) \right). \quad (1.11)$$

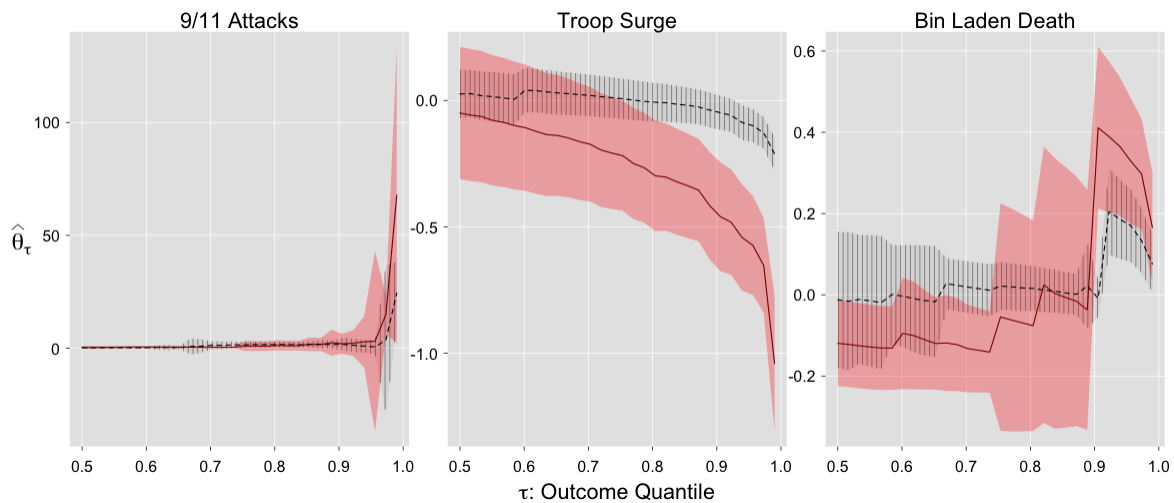
In words, for each partition of the data  $\pi_{\tau}$ , we solve for the linear function  $\mu(\nabla_i, \theta)$  which provides conditional estimates for the  $\tau^{\text{th}}$  quantile.

Results of this analysis are provided in Figure 1.5, which provides results over two different measures of  $\overline{\text{AR}}_i$ . Baseline BSTS estimates utilized to derive coefficient estimates. Kernel standard errors are provided around each estimate. The estimates marked (---) use a post-event window length of  $T - T_0 = 100$  to calculate  $\overline{\text{AR}}_i$ , while the estimates marked (—) use a post-window length of  $T - T_0 = 20$ . Firm-level abnormal returns were

<sup>21</sup>An analysis that compares estimates of abnormal returns with product-level dependencies can be found in Appendix A.15.



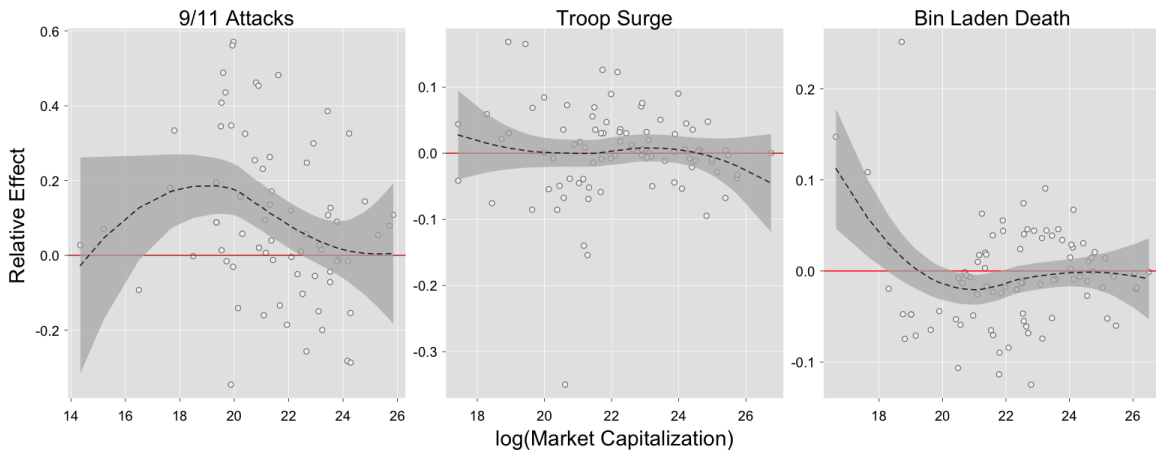
**Figure 1.5:** How Firm-Level Abnormal Returns Map to Quantiles of DoD Revenue Dependence



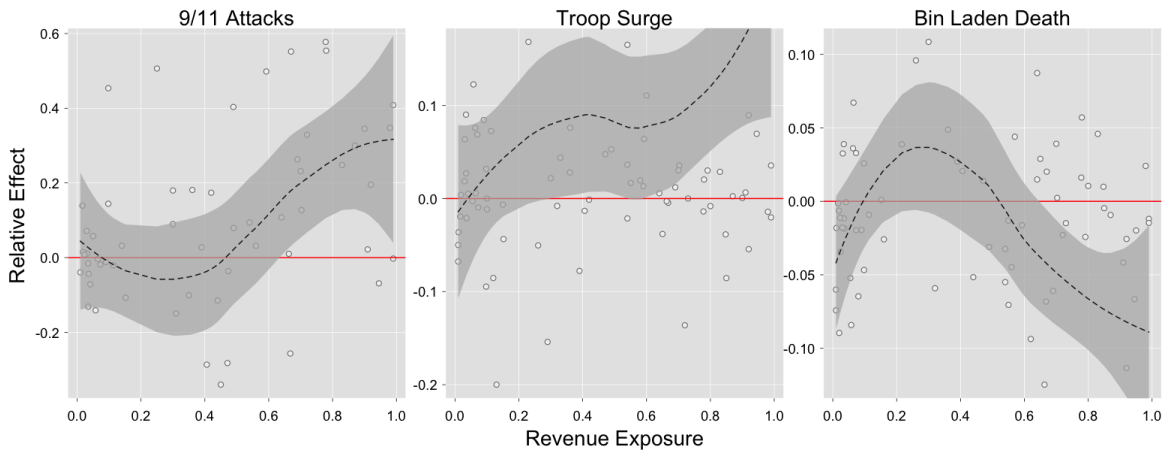
*Notes:* Estimates marked (—) use a post-window length of  $T - T_0 = 20$ , while estimates marked (---) use a post-event window length of  $T - T_0 = 100$ .  $\hat{\theta}_\tau$  can be interpreted as a quantile-specific linear regression coefficient between revenue dependence and abnormal returns, across quantiles of abnormal returns. In the 9/11 and Bin Laden Death event studies, there is a positive association across firms between the top decile of abnormal returns and revenue dependence. A differing picture is shown in the Iraq Troop Surge study, where higher returns tended to come from firms with lower levels of revenue dependence. These results suggest the revenue dependence measure has unequal predictive power across events, and that differing events may impact qualitatively differing sets of firms. Just over 400 firm-level estimates (i.e., the number of publicly-traded exposed firms with at least 8 quarters of prior SEC data) were utilized to create the graphic for the 9/11 event study; over 800 firms each were utilized to construct the Iraq Troop Surge and Bin Laden Death graphics.

**Figure 1.6: Abnormal Returns versus Company Characteristics (High-Earning Subset of Firms)**

*(a) Abnormal Returns versus Company Size*



*(b) Abnormal Returns versus Revenue Dependence to DoD*



*Notes:* This figure plots the relationship between a firm's market capitalization (on the last trading day before each event) against its estimated relative effect, given by  $\text{AR}_i$  with  $T - T_0 = 20$ . The curved line in each subplot is a local regression estimate of the relationship with 95% intervals.

standardized—i.e., demeaned and divided by the sample standard deviation—prior to quantile regression. Across all three event studies we see that the conditional dependence of  $\overline{AR}_i$  with respect to  $\nabla_i$  appears to be nonlinear. Across events, firms with “more extreme” financial returns were more likely to have higher levels of revenue dependence.

Motivated by this observation, Figure 1.6 plots the results of the event studies for a relevant subset of “high-earning” firms (i.e., public firms that were at any point in the sample among the top 100 grossing defense firms in a given financial year). The top panel of the graphic plots the relationship between firm-level relative effects ( $\overline{AR}_i$  estimates with  $T - T_0 = 20$ ) and the logarithm of each firm’s pre-event market cap. Curved lines in each subplot mark local regression lines with 95% confidence intervals. In general, larger firms do not appear to be systematically related to abnormal returns—a pattern which holds true across event studies. The lower panel of Figure 1.6 tells a different story, however, as it plots abnormal returns against firm-level measures of revenue dependence. The relationship between revenue dependence and abnormal returns appears positive in the 9/11 and Troop surge event studies, but the relationship is negative in the Bin Laden event studies. Amongst high-earning firms, firms with the higher levels of revenue dependence to the DoD tend to be those that respond most extremely across events, but there is still considerable variation left unexplained.

### 1.4.3 In-Sample Performance of Competing Statistical Approaches

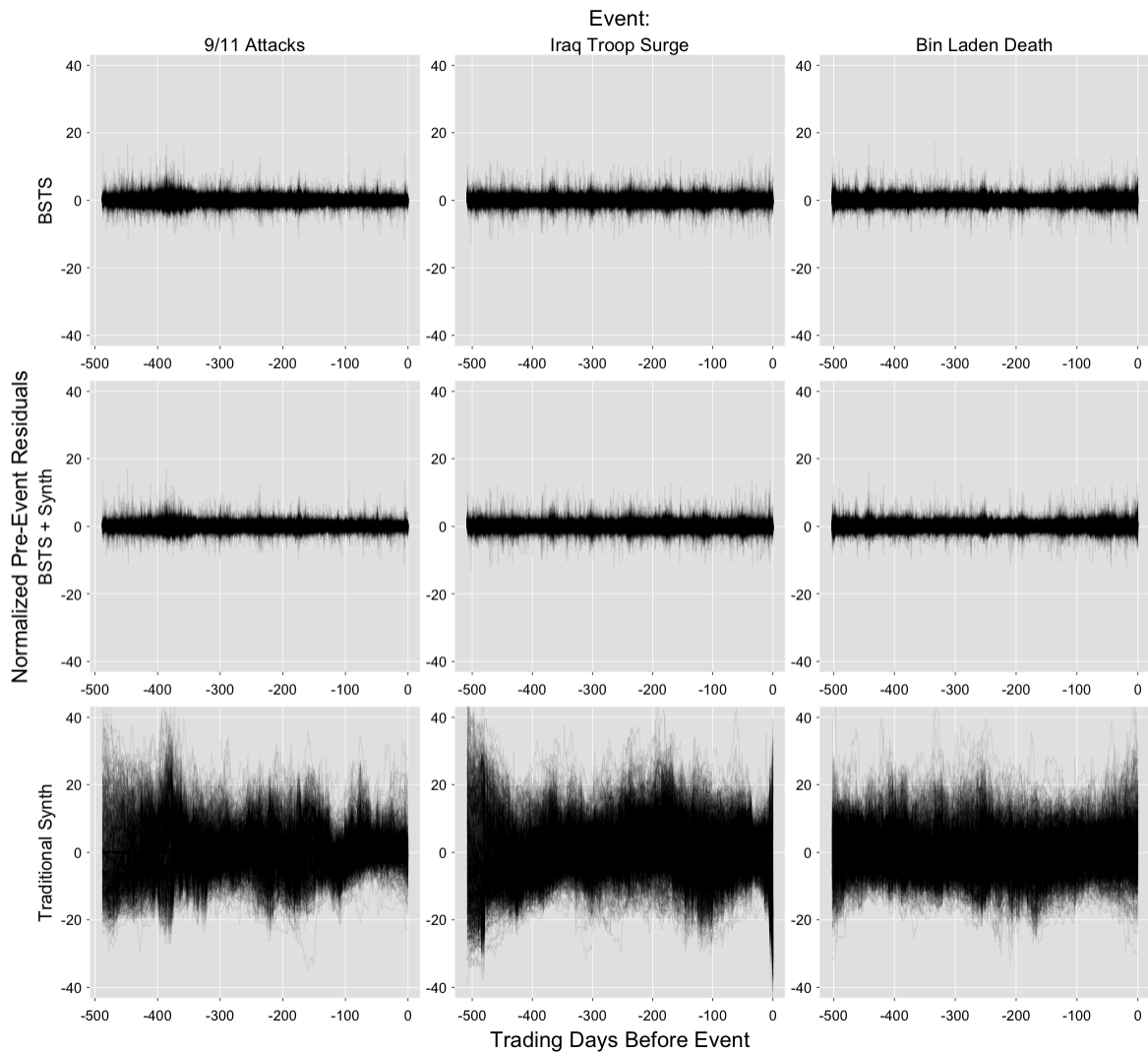
Results presented in this section suggest that firm valuations detectably respond to stimuli such as those analyzed, and the degree of firm financial response to individual events is moderated by a firm’s degree of financial dependence on the DoD. However, a natural question remains: how well does the BSTS perform in this context relative to competing approaches, such as the synthetic control estimator? Figure 1.7 provides motivating evidence for the claim that the BSTS outperforms the traditional synthetic control in our setting. The 3x3 grid structure segments competing estimators on the

horizontal axes against events labeled on the vertical axes. Within each subplot, thin black lines represent measures of normalized residuals in the pre-event window. More precisely, within a given event study, the “normalized pre-event residuals” for firm  $i$  given estimator  $l$  are provided by  $(Y_{it} - \widehat{Y}_{it}^l) / \left( \sqrt{\frac{1}{T_0-1} \sum_{t=1}^{T_0} (Y_{it} - \widehat{Y}_{it}^{bsts})^2} \right)$ . This is analogous to a z-score transformation where each firm’s residuals are divided by the standard deviation of the baseline BSTS model residuals. The graphic reveals that the BSTS models generally outperform the traditional synthetic control estimator in terms of pre-event fit. The BSTS models exhibit considerably lower autocorrelation in residuals, and point-wise residuals from the traditional synthetic control estimator are routinely 10 times the size of those from the BSTS models. Given that Abadie, Diamond and Hainmueller (2010, 2015) indicate a worthwhile “sanity check” for evaluating a synthetic control’s appropriateness comes through one’s inspection of its pre-event residual series, the fact that BSTS models consistently outperform the traditional synthetic control estimator by this measure is worth noting.<sup>22</sup> Given that our causal inferences are drawn using observational data, however, one might be concerned that ostensibly superior pre-event residual series are simply the byproduct of a model’s overfitting. If we could (in an ideal case) perfectly observe firm-level counterfactual series, we would compare results derived from the BSTS against those derived from the synthetic control estimator. To this end, the next section of this paper introduces a set of simulations (i.e., “Monte Carlo” studies) to assess the relative performance of the BSTS and synthetic control paradigms. Overall, results of the simulation studies corroborate findings introduced in this subsection. While the BSTS and synthetic control methods both recover correct causal estimates in typical conditions, the BSTS estimates routinely have lower variance.

---

<sup>22</sup>Analogous series may be plotted by looking at the degree of variation in the post-event window. Overall, causal estimates from the traditional synthetic control estimator exhibit a far greater degree of period-to-period variation in the post-event window as well. These findings additionally suggest the synthetic control estimator routinely underperforms in this setting.

**Figure 1.7:** *Relative Fit of Counterfactual Estimates in Pre-Event Window (Scaled to Baseline BSTS Model)*



*Notes:* Each subplot visualizes the relative fit of a particular estimator in the pre-event window relative to the baseline BSTS model. In each subplot there are approximately 1,000 individual time series plotted—one for each “exposed” firm in the event study. The varied color gradient in each subplot is a function of the density (or overlap) of plotted series.

## 1.5 Monte Carlo Studies to Assess Model Performance

This paper makes the argument that BSTS is an appropriate methodology for estimating causal effects in dynamic settings such as ours. Section 1.4 demonstrates that the BSTS tends to outperform the traditional synthetic control estimator in terms of model fit in the pre-event window, although the inclusion of the Abadie, Diamond and Hainmueller (2010, 2015) estimate as a potential control variable weakly improves upon the BSTS model that relies upon aggregate market indices. With observational data alone, however, judgments about the suitability of the BSTS estimator are limited by the fact that one never truly observes individual treatment effects.

To assess how well the BSTS performs under a range of hypothetical data generation processes, this section of the paper performs Monte Carlo studies of simulated financial time series. In each of these studies, the true value of the simulated treatment effect for a given firm  $i$  is  $\phi_i = 5$ , and the length of the pre- and post-event windows are 500 and 100, respectively. To assess the conditions under which the BSTS performs relatively better or worse, we repeatedly regenerate random data under a varied set of structural conditions and store the results of the model in each iteration. In the first study (Section 1.5.1), autoregressive financial series are simulated with random correlations between the exposed series and potential control variables, where levels of autocorrelation in a firm's moving average and period-specific errors are motivated by observed patterns in financial data. In the second study (Section 1.5.2), financial series are assumed to be generated from unobserved, latent series that are determined by independent random walks, while firms randomly vary in terms of their own unit-level random walks and their financial dependence on latent factors. Principal components analysis is performed on these simulated data to approximate the case in which a firm's financial wellbeing may be a function of numerous underlying (independent) dimensions, firm's have individual-level esoteric drift, errors are autocorrelated, and despite the fact that an exposed series may appear only weakly correlated with an individual factor loading. We then compare

the performance of the BSTS model against the traditional synthetic control estimator.

Results of the simulation studies reveal the BSTS effectively recovers causal effects across these settings. Overall, three main conclusions arise from these studies: (a) the BSTS estimator appears to be “unbiased” across simulations, (b) the precision of the estimator depends on the “informativeness” of the potential covariates, and (c) unlike the traditional synthetic control estimator, the model performs equally well when associations between the treated and control series are negative. In the space below we describe in detail the setup of these simulation studies and their respective results.

### 1.5.1 Simulation Study 1: Random Correlations between Exposed Series and Controls

#### Setup

Financial time series are known to exhibit high-levels of autocorrelation, just as daily returns are known to be correlated across the market. Consider Table 1.1, which marks pair-wise correlations between prominent assets over the two years of trading data prior to September 11, 2001. The upper panel of Table 1.1 is a correlation matrix of individual share prices (i.e., for any row  $i$  and column  $j$ , the cell maps  $\text{Corr}(Y_{it}, Y_{jt})$ ); the lower panel denotes correlations between daily returns (i.e.,  $\text{Corr}([Y_{it} - Y_{i,t-1}]/Y_{i,t-1}, [Y_{jt} - Y_{j,t-1}]/Y_{j,t-1})$ ). Quite clearly, we observe there are associations between share prices within a given sector, but also correlations in the cross-section in terms of daily returns. But firm valuations are not simply correlated within sectors over time, as we see that within-firm daily returns are also autocorrelated. Consider Figure 1.8, which plots the autocorrelation function (ACF) and partial autocorrelation function (PACF) of Northrop Grumman’s share price over time. Over the two years of trading data, the average correlation between Northrop Grumman’s share price in levels in time  $t$  and time  $t + 1$  is about 0.99. The PACF plot reveals that conditional on a prior periods’ valuations, there is

a strong association between the firm's share price in adjacent periods; but conditional on the lag-1 value, the association between Northrop Grumman's share price and prior periods' prices are very weak, near zero, and within acceptable bounds of what would be expected from a stationary process. Substantively similar takeaways come from looking at the ACF and PACF plots for daily returns rather than share prices.

**Table 1.1:** Correlations of Daily Prices and Returns for Prominent Firms (Two Years of Trading Data Prior to 9/11)

*Correlation Matrix of Daily Valuations:  $Y_{it}$*

	SP500	LMT	BA	NOC	GD	GE	MSFT	APPL	AMZN
SP500	1	-0.70	-0.47	-0.48	-0.66	0.64	0.42	0.76	0.59
LMT	-0.70	1	0.82	0.89	0.94	-0.07	-0.75	-0.85	-0.87
BA	-0.47	0.82	1	0.82	0.83	0.20	-0.69	-0.79	-0.78
NOC	-0.48	0.89	0.82	1	0.82	0.21	-0.88	-0.72	-0.89
GD	-0.66	0.94	0.83	0.82	1	-0.01	-0.70	-0.83	-0.84
GE	0.64	-0.07	0.20	0.21	-0.01	1	-0.16	0.26	-0.06
MSFT	0.42	-0.75	-0.69	-0.88	-0.70	-0.16	1	0.66	0.84
APPL	0.76	-0.85	-0.79	-0.72	-0.83	0.26	0.66	1	0.73
AMZN	0.59	-0.87	-0.78	-0.89	-0.84	-0.06	0.84	0.73	1

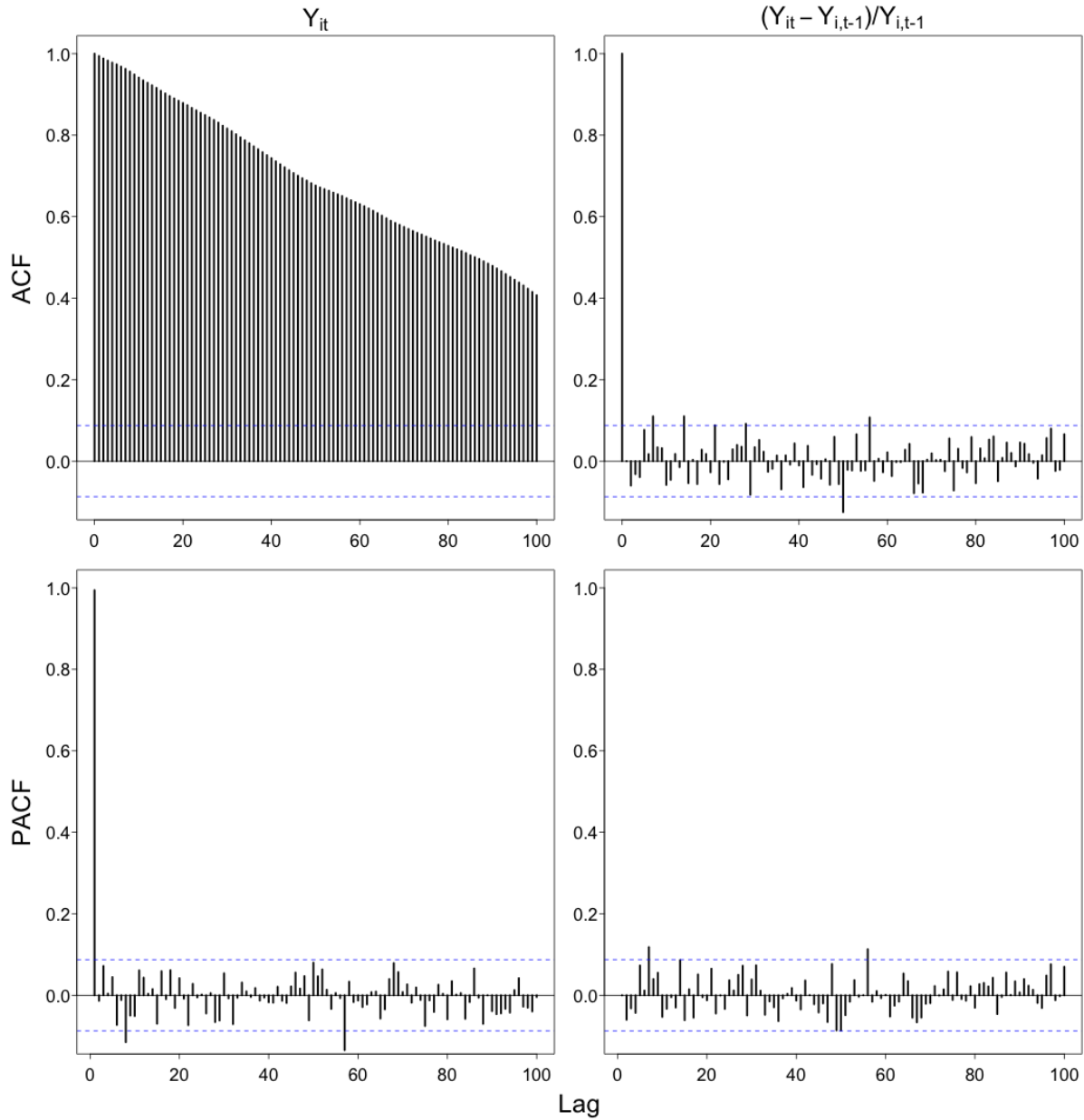
*Correlation Matrix of Daily Returns:  $(Y_{it} - Y_{i,t-1})/Y_{i,t-1}$*

	SP500	LMT	BA	NOC	GD	GE	MSFT	APPL	AMZN
SP500	1	0.13	0.32	0.24	0.31	0.68	0.57	0.47	0.42
LMT	0.13	1	0.24	0.37	0.33	0.06	-0.05	-0.07	-0.02
BA	0.32	0.24	1	0.29	0.37	0.26	0.10	0.07	0.07
NOC	0.24	0.37	0.29	1	0.39	0.11	0.03	0.06	-0.02
GD	0.31	0.33	0.37	0.39	1	0.23	0.10	0.10	0.01
GE	0.68	0.06	0.26	0.11	0.23	1	0.30	0.28	0.20
MSFT	0.57	-0.05	0.10	0.03	0.10	0.30	1	0.37	0.33
APPL	0.47	-0.07	0.07	0.06	0.10	0.28	0.37	1	0.28
AMZN	0.42	-0.02	0.07	-0.02	0.01	0.20	0.33	0.28	1

*Notes:* Stock trading symbols correspond to SP500 for the S&P500 market index, LMT for Lockheed Martin, BA for Boeing Airlines, NOC for Northrop Grumman, GD for General Dynamics, GE for General Electric, APPL for Apple Computers, and AMZN for Amazon.com. A sample of prominent defense firms and technological firms are chosen to highlight the association between financial time series within sectors and across sectors. Daily price levels are fairly strongly correlated within sectors. For example, the correlation between LMT and NOC's daily company valuations was 0.89 in the two years of data prior to 9/11. Daily returns are also more strongly correlated within sectors than across sectors. Daily returns are positively associated with market-wide returns on average, though firm-level associations with market returns vary to some degree.



**Figure 1.8:** Autocorrelation of Stocks in Sample (Example: Northrop Grumman Daily Share Price)



Notes:  $Y_{it}$  is Northrop Grumman's valuation in time  $t$ , and  $(Y_{it} - Y_{i,t-1})/Y_{i,t-1}$  is its daily return. The rows marked ACF and PACF indicate the autocorrelation function and partial autocorrelation functions, respectively. The horizontal axis plots time. Overall the  $Y_{it}$  exhibits a high degree of autocorrelation—a pattern that is not unique to Northrop Grumman. The diagrams suggest the conditional independence of returns can be fairly well-approximated by an AR(1) series, as lagged correlations in the ACF and PACF are within acceptable bounds. This observation motivates the lag structure utilized in the main analysis and simulation studies.

To test the validity of the BSTS approach on data generated with this sort of dependence, we generate financial time series with the following structure:

$$y_{it} = a_{it} + e_{it} \quad (1.12)$$

where  $a_{it} = \rho_i \cdot a_{i,t-1} + u_{it}$ ,  $\mathbf{u} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ,  $e_{it} = r \cdot e_{i,t-1} + v_{it}$ , and  $v_{it} \sim_{iid} N(0, 1)$ . In words, the value of firm  $i$  in period  $t$  is given by the firm-level autoregressive series  $\{a_{it}\}$  with associations to other firm-level series given by the covariance matrix  $\boldsymbol{\Sigma}$ , plus a firm-specific autocorrelated error. Recall that for any two firms  $i$  and  $j$  with series generated as in Equation 1.12, the expected value of the correlation between any two such autoregressive series is

$$\mathbb{E}(\text{Corr}(y_{it}, y_{jt})) = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \cdot \frac{1 - \rho_i\rho_j}{\sqrt{(1 - \rho_i^2)(1 - \rho_j^2)}} \quad (1.13)$$

where  $\sigma_{ii}$  is simply the variance of the autoregressive series  $y_{it} = a_{it} + e_{it}$ , and  $\sigma_{ij}$  the covariance between the respective series. From this we see that given a either a positive-definite covariance matrix  $\boldsymbol{\Sigma}$  or its equivalent correlation matrix, we can simulate autoregressive processes of a desired correlation in expectation.

## Implementation and Results

Assume there is one exposed series and  $J = 10$  control series. There are  $T_0 = 500$  pre-event periods and 100 post-event periods, for a total of  $T = 600$  time periods. In line with Equations 1.12 and 1.13, for a given simulation  $s \in 1, \dots, 5000$ , we generate series

with the structure

$$\mathbf{y}_t = \begin{pmatrix} y_{1t} = D \cdot \phi + a_{1t} + e_{1t} \\ y_{2t} = a_{2t} + e_{2t} \\ \vdots \\ y_{J+1,t} = a_{J+1,t} + e_{J+1,t} \end{pmatrix}, \text{ where } D = 1 \text{ if } t > T_0, \text{ and } D = 0 \text{ otherwise,} \quad (1.14)$$

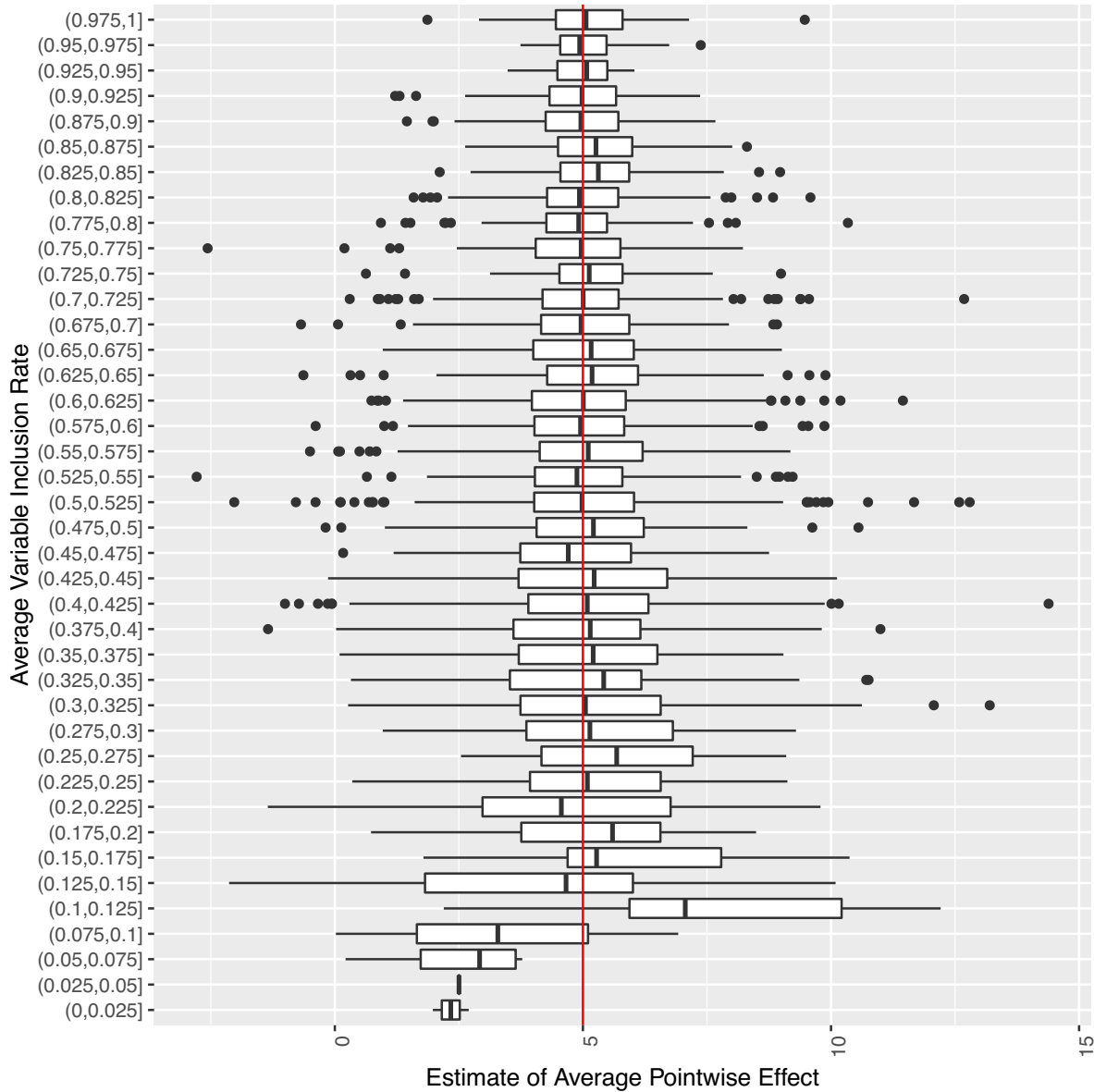
with a randomly generated positive-definite covariance matrix  $\Sigma^{\{s\}}$  fixed throughout simulation  $s$ , and as randomly generated by the approach outlined in Joe (2006). Across simulations the true effect of exposure is set to  $\phi = 5$ , to approximate a treatment effect size in the small to medium tier.<sup>23</sup> We set  $\rho_i = 0.99$  and  $r_i = 0.1$  for all  $i$  across all simulations. Within each simulation  $s$  we estimate the BSTS model as outlined in Equation C.5 from Section 1.3, and store the estimate  $\hat{\phi}^{\{s\}}$ . We are interested in (a) the relationship between  $\hat{\phi}^{\{s\}}$  and the true  $\phi$ , and (b) the conditions under which the BSTS performs relatively better or worse as a function of potential covariates included.

Results of this first simulation study are presented in Figure 1.9. Overall we see that as average variable inclusion rates increase, the variance of the model estimates decreases substantially. Furthermore, the vertical line in the figure marks the true treatment effect of  $\phi = 5$ . Aggregating over all 5,000 independent simulations, the mean estimated effect was  $\frac{1}{5000} \sum_s \hat{\phi}^{\{s\}} \approx 5.04$ , which is indistinguishable from the true value of  $\phi = 5$  given the sample size. This result suggests that BSTS model performance is largely determined by the predictive power of the inputs used as covariates. In the event of a weak relationship between potential control series and the exposed series, the model estimates are less efficient.

---

<sup>23</sup>This value was chosen because the standard deviation of a typical defense company's daily return is between 2-2.5%. A treatment effect of  $\phi = 5$  is therefore meant to approximate the magnitude unusual daily return, but is far lower in magnitude than estimated abnormal returns for major firms after 9/11 or the troop surge, for example.

**Figure 1.9:** *BSTS Performance versus Variable Inclusion Rates with Data from Random Covariance Matrices (Study 1)*



*Notes:* This figure plots estimates from Study 1, where the BSTS is estimated using exposed series and controls generated from random, positive-definite covariance matrices (Joe, 2006). The horizontal axis plots the average variable inclusion rate from the “spike-and-slab” procedure outlined in Section 1.3. Each boxplot in each row maps the empirical distribution of effects estimated within a given threshold marked on the horizontal axis, where intervals on the horizontal axis indicate average variable inclusion rates from a given simulation  $s$ . For example, the line marked  $(0.975, 1]$  provides a box plot of the distribution of model estimates for BSTS models with average variable inclusion rates between 0.975 and 1. Within each boxplot, solid-vertical lines denote the median, the edges of the boxes denote the 25th and 75th percentiles, and the range of the thin-horizontal lines mark the 5th and 95th percentiles.

## 1.5.2 Simulation Study 2: A Multi-factor Model with Random Dependence on Principal Components

### Setup

Assume there are  $M > 0$  total latent “factors” in the economy, where  $m$  denotes an individual factor. One such dimension represents the valuation of an orthogonal sector of the economy, whereby individual firm valuations are a function of the performance of these sectors. Our goal is to see how well the BSTS estimator performs with principal components estimated from a sample of control firms. To do this, we must simulate financial time series where

$$a_{it} = \sum_{\tau=1}^t a_{i\tau}, \quad a_{i\tau} \sim_{iid} N(0, \sigma_i), \quad \text{for } \tau > 0, \quad (1.15)$$

$$b_{mt} = \sum_{\tau=1}^t b_{m\tau}, \quad b_{m\tau} \sim_{iid} N(0, \sigma_m), \quad \text{for } \tau > 0, \quad (1.16)$$

$$u_{it} = \rho_i \cdot u_{i,t-1} + \epsilon_{it}, \quad \epsilon_{it} \sim_{iid} N(0, \sigma_\epsilon), \quad \text{for } t > 0, \quad (1.17)$$

$$w_{im} = \frac{(\theta_{im})^q}{\sum_k (\theta_{im})^q}, \quad \theta_{im} \sim_{iid} U(0, 1), \quad (1.18)$$

where  $a_{i0} = 0$ ,  $b_{m0} = 0$ , and  $u_{i0} \sim_{iid} N(0, \sigma_\epsilon)$ . A firm’s valuation in a given time period is given by:

$$y_{it} = a_{it} + \sum_m (w_{im} \cdot b_{mt}) + u_{it}, \quad (1.19)$$

As such, firm valuations are determined by unit-level, uncorrelated random walks, esoteric levels of sector-specific and market-wide influence, and unit-level autocorrelated errors. Notice the exponent  $q$  used to determine  $w_{im}$  may be thought of as a quantity that scales the degree to which firms specialize more heavily in one sector as opposed to another. Note that as  $q$  approaches zero, the quantity  $w_{im}$  converges to  $1/M$ , or one over the total number of sectors in the market. In this world, firm valuations are effectively driven by a firm-level random walk, a market-wide effect (which may be thought of as analogous to the case in which  $M = 1$ , but with a per-period variance inflated by the

factor  $M$ ), and firm-level autocorrelated errors. When  $q = 1$ , the expected value of a firm's maximum sector weight is equal to

$$\begin{aligned} \mathbb{E}(\max\{w_{im}\}_{m=1}^M) &= \frac{\mathbb{E}(\max\{\theta_{im}\}_{m=1}^M)}{\sum_m \mathbb{E}(\theta_{im})} \\ &= \frac{\int_0^1 M \cdot \theta^{M-1} d\theta}{M \cdot \mathbb{E}(\theta_{im})} \\ &= \frac{M/(M+1)}{M/2} = \frac{2}{M+1}. \end{aligned} \quad (1.20)$$

When  $q > 0$ , firms specialize much more uniquely in individual sectors of the economy, and their market valuations are more heavily dependent on unique sectors' random walks.

### Implementation and Results

We simulate 5,000 independent economies using the structure outlined in Equation 1.19, each with 251 individual firms. We vary  $q \in \{0.1, 0.5, 1, 2.5\}$  across fourths of these simulations, though we set  $M = 10$  in each. Values of  $b_{mt}$  are common across firms in an economy, while  $a_{it}$ ,  $u_{it}$ , and  $w_{im}$  vary at the firm level. In each simulation we assume  $\rho_i = 0.1$ ,  $\sigma_\epsilon = 0.5$ ,  $\sigma_m = 1$ , and  $\sigma_i = 0.1$ . These values are intended to mimic observed levels of autocorrelation in individual firm returns over time (e.g., as shown in Table 1.1 and Figure 1.8), in addition to the fact that a firm's daily returns tend to be more strongly correlated with market-wide effects than their own lagged outcomes.

In each simulation, we assume without loss of generality the first firm is "exposed" to the treatment and the remaining 250 firms are potential control firms. The exposed firm's outcomes are given by

$$y_{1t} = \begin{cases} a_{1t} + \sum_m (w_{1m} \cdot b_{mt}) + u_{1t} & \text{if } 1 < t \leq T_0 \\ \phi + a_{1t} + \sum_m (w_{1m} \cdot b_{mt}) + u_{1t} & \text{if } T_0 + 1 < t \leq T \end{cases} \quad (1.21)$$

where  $\phi = 5$ ,  $T_0 = 500$  and  $T = 600$ . In each simulation we perform principal component

analysis on the 1,000 control firms, keeping the only the first 10 factor loadings (just as presented in the main body of the text), and we estimate the BSTS utilizing the first 10 principal components as potential covariates, taking 2,500 MCMC draws in each simulation. Results of this analysis are presented in Table 1.2. Across all simulations we see that the BSTS recovers the true treatment effect on average, as marked in the column labeled “Mean  $\hat{\phi}$ .” Overall, the BSTS provides unbiased causal estimates of the effect of the treatment, and the level of  $q$  does not appear to associated with the validity of estimates from the model. This gives us great confidence in the methodological approach for estimating causal quantities generated from processes such as Equation 1.19.

To assess how well the BSTS performs relative to the traditional synthetic control estimator, we repeat the above process but generate traditional synthetic controls in each simulation. More precisely, we generate exposed and control series as above—but rather than simply use the PCA factor loadings as potential covariates, we also include the estimate provided by Abadie, Diamond and Hainmueller (2010, 2015) as a possible model covariate. Results of this analysis are provided in Figure 1.10. The top panel of this graphic shows the approximate bias of the composite BSTS model (i.e., the model that includes synthetic controls as a potential covariate) versus the bias from the traditional estimator. Points are colored according to their local kernel density, with warmer colors indicating a greater density of points in that area. Overall we see that both methods appear to recover the true causal effect on average, but estimates from the traditional synthetic control model exhibit greater variance. The standard deviation of estimates from the traditional synthetic control estimator is 4.22, while the associated standard deviation from the BSTS is 2.12.

However, to what degree does the inclusion of the synthetic control as a potential control variable in the BSTS improve over the baseline model? The lower panel of

**Table 1.2:** BSTS Performance in Principal Components Experiment (Study 2)

$q$	Mean $\hat{\phi}$	SD	5%	25%	50%	75%	95%	Sims
0.1	4.89	2.56	0.55	3.78	5.00	6.15	8.55	1250
0.5	5.17	2.41	1.45	4.00	5.10	6.34	9.06	1250
1	5.01	2.48	1.06	3.88	5.02	6.14	8.96	1250
2.5	4.93	2.43	0.92	3.77	4.98	6.14	8.57	1250
Total	5.00	2.47	1.04	3.85	5.03	6.19	8.84	5000

*Notes:* Across simulations, the true treatment effect is  $\phi = 5$ . Across the 5000 simulations, the average estimate of this effect is almost exactly 5. Columns with percentage signs mark quantiles of estimated effects. Notably, the model appears to perform equally well across levels of  $q$ . Quantiles of estimated effects reveal estimates are well-behaved and normally distributed about the mean estimate.

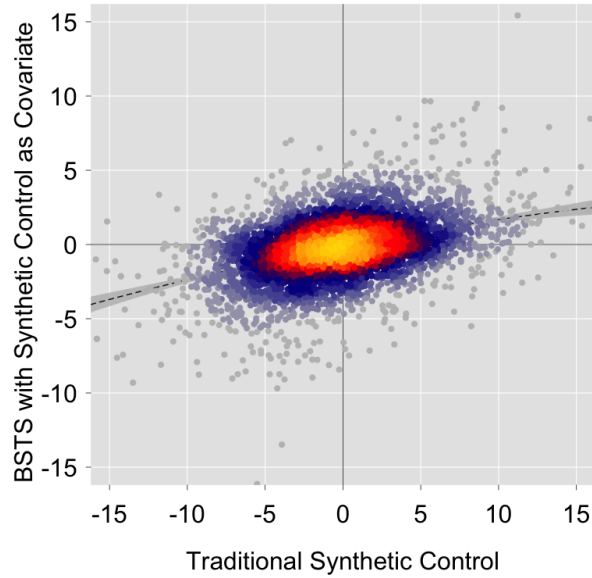
Figure 1.10 inspects this question, as it plots the relative absolute error<sup>24</sup> of a given estimator—benchmarked to the traditional synthetic control estimator—across quantiles of each statistic. Overall, we see that the inclusion of the traditional synthetic control appears to weakly improve over the baseline BSTS model, but distinct performance gains appear only to have been present in about 10% of simulations. The BSTS and composite models consistently exhibit lower errors relative to the traditional synthetic control, and that relation holds constant across nearly all simulations. In the minority of simulations where the relative error of the baseline BSTS was greater than that from the synthetic control, the composite model still outperformed the baseline BSTS and the traditional synthetic control estimator.

<sup>24</sup>In this case, the absolute error of a given model  $l$  within a given simulation  $s$  is equal to  $AE_l = |\phi - \hat{\phi}_l^{\{s\}}|$ . Let  $AE_{l\tau}$  denote a given model's absolute error at quantile  $\tau$ . The relative absolute error for model  $l$  at quantile  $\tau$  is given by  $RAE_{l\tau} = \frac{AE_{l\tau}}{AE_{Synth,\tau}}$

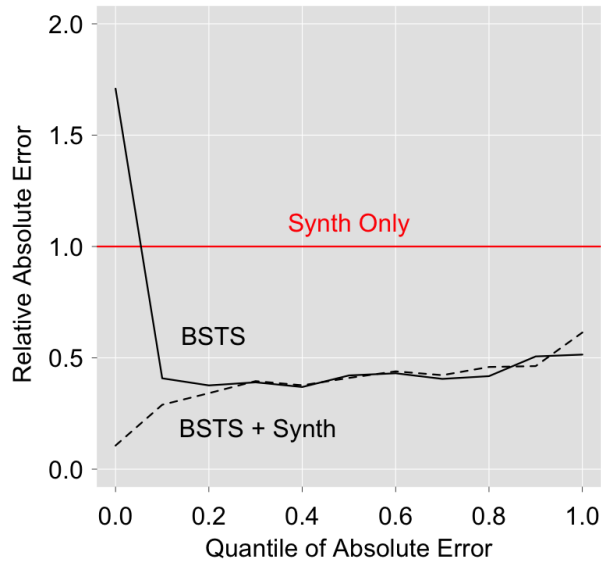


**Figure 1.10:** Performance of Competing Estimators (Study 2)

(a) Both Unbiased, BSTS Has Lower Variance



(b) Estimators' Relative Absolute Errors



Notes: The top panel graphs the simulation-level deviation (or “bias”) of the traditional synthetic control estimator against the composite BSTS model. Simulation-level biases are for a given estimator  $l$  are given by  $\phi - \hat{\phi}_l^{\{s\}} = 5 - \hat{\phi}_l^{\{s\}}$ . In general, there is a positive association between estimates across the models, as indicated by the local regression line in the background of the graphic. The lower panel marks the relative absolute error of a given estimator across quantiles of all simulations. Hence, the relative absolute error for the traditional synthetic control estimator is equal to 1.

## 1.6 Discussion

A central tenet in international relations theory is that war is costly and, all else equal, government actors should prefer peacetime to war (e.g., Fearon, 1995). To prepare for and participate in wars places enormous social and economic costs on host nations (e.g., Belasco, 2014; Stiglitz and Bilmes, 2008), threatens prospects for longterm growth and foreign investment (e.g., Abadie and Gardeazabal, 2003; Abadie, Diamond and Hainmueller, 2015; Collier, 1999; Deger and Sen, 1995; Mintz and Huang, 1992; Ram, 1995), diverts resources away from other sectors of the economy (e.g., Mintz, 1989; Powell, 1993; Ramey and Shapiro, 1999), and may influence a nation's monetary policy in the short term. In the years since 9/11, social scientists have had an understandable and resurgent interest in understanding both the economic consequences of conflict and the political phenomena that underpin decisions fund and fight in wars (e.g., Arena and Wolford, 2012; Phillips, 2015; Whitten and Williams, 2011; Williams, 2015).<sup>25</sup> Understanding the costs of war has been for decades, and will continue to be, a major focus of scholarly research on conflict processes.

With all this said, researchers know close to nothing about how individual actors actually benefit (or are harmed by) sudden shifts in the international security environment, the demand for defense, how the defense industry financially responds—in an empirical sense—to the dynamics of ongoing conflicts, much less how individual stakeholders may be helped or hurt by the dynamics of international conflict. While the field of international relations has had a renewed interest in the behavior of “non-state” actors in recent years, there is little quantitative research that links the valuation or behavior of these interest groups to shifts in conflict or violence. The lack of scholarship on this subject is concerning but to be expected. It is concerning because of the sheer magnitude of the dollars spent in the defense industry domestically and globally. It is perhaps to

---

<sup>25</sup>See Sandler (2014) for a worthwhile overview of the types of questions researchers commonly ask in this literature.

be expected, however, due to the challenge of working with troves of company-level and spending data, aligning those data with military events, and finding a plausible identification strategy for estimating individual-level causal effects.

This paper has attempted to take a necessary first step in understanding to what extent individual defense firms profit from unexpected shifts in the international security environment and documenting the considerable financial heterogeneity within the defense sector. The analysis makes headway on a remarkably understudied area of political life, despite the ever-growing need to understand the ramifications of defense budgeting in our world today. The consequences of defense policy extend far beyond the direct military services they provide, after all. Military spending accounts for roughly two to three percent of global GDP, or about 2 trillion dollars per year since FY2000. Over that same time period, U.S. expenditures alone account for approximately 40% of global military spending. But defense spending is not simply important because of the total dollars spent on such services, but also due to its volatility. In the United States, roughly ninety percent of the year-to-year variation in the Federal budget is explained by shifts in the defense budget (Ramey, 2011). Relative to all other discretionary programs, the U.S. Department of Defense commands a greater share of the budget than the Departments of Treasury, Education, Health, State, and all other departments combined. As a result, shifts in expectations over defense spending—which tend to increase in times of heightened security concerns—are a major source of macroeconomic volatility, which research suggests can have a profound (if not distortionary) effect on global wellbeing.<sup>26</sup>

Results of the analysis suggest there is considerable variation in how individual firms respond to major shifts in expectations over U.S. foreign policy. While estimated firm level effects (in levels) are associated with overall DoD outlays, estimated abnormal returns (in

---

<sup>26</sup>Different literatures on these themes focus on how shifts in defense spending may influence economies (e.g., Barro, 1981; Caplan, 2002; Mintz and Huang, 1992; Ram, 1995; Ramey and Shapiro, 1999; Reich, 1972; Schneider and Troeger, 2006), or how the drawdown of defense purchases may cause recessions (Phelan and Trejos, 2000), for example.

percentage point terms) are more strongly correlated with how revenue dependent an individual firm is the Department of Defense. Amongst a subset of high-earning defense firms, surprise shocks in the international security environment can lead individual firms to over-perform or under-perform their expected valuations by billions of dollars over a relatively short time horizon. These findings not only suggest that prominent events shift expectations over the demand for defense, but the news of shifting expectations is quickly priced into financial markets. Furthermore, not all major defense firms react uniformly to these events. Firms that specialize in giving healthcare to soldiers and veterans (e.g., Health Net), for example, may behave in financial markets in a fashion similar to major weapons producers (e.g., Raytheon) after such events; but simply because two firms yield similar aggregate contract dollars and produce similar goods does not mean they will respond to the same stimulus in the same way (e.g., Boeing and Lockheed Martin, which since FY2000 have been obligated roughly \$300 Billion USD and \$400 Billion USD, respectively).

To better understand how individual firms react to external stimuli requires understanding the full revenue streams that matter to individual firms, accounting both for government contracts and non-public dollars. To this end, this paper tabulated information from SEC financial statements and merged aggregate company revenues with official Department of Defense procurement data. This new dataset permits firms to be scaled by the degree to which DoD contract contract dollars matter to their aggregate revenue streams. Overall, firms that are more revenue dependent on the DoD respond most “abnormally” following major international security events. Across the three events studied, revenue dependence appears to be a far stronger predictor of abnormal returns than the individual products or services that a firm specializes in.<sup>27</sup> One possible explanation for this result is that many prominent defense firms, such as Lockheed Martin

---

<sup>27</sup>Overall there is a positive association between portfolio “diversity,” as measured by total number of unique services provided by a firm to the DoD, and overall contract dollars. Supplemental analyses on this point are present in the Appendix.

and Northrop Grumman, provide hundreds of different products to the Department of Defense in a given year, while other major contractors, such as Health Net and Oshkosh Corporation, specialize much more specifically in a single class of products. Nevertheless, a worthwhile direction for future research would be to investigate more rigorously the relationship between product-level revenue streams and abnormal financial behavior. Such an analysis, however, falls outside the scope of this particular paper.

But beyond its substantive focus, this paper has made headway on methodological issues that extend beyond the applied question at hand. A chief methodological aim in the paper has been to compare the relative merits of the influential Abadie, Diamond and Hainmueller (2010, 2015) synthetic control estimator against the Bayesian structural time series approach first introduced by Varian (2014) and Brodersen et al. (2015). As applied to the study of financial time series, this paper makes the argument that strategies such as the BSTS appear to be superior to the traditional synthetic control approach across a range of dynamic settings. In the rare case the traditional synthetic control is a better approximation of an exposed unit's counterfactual series than a baseline BSTS model, the paper shows that one may nest the traditional synthetic control estimate within the BSTS framework to provide causal estimates that on average outperform both the traditional synthetic control and BSTS approaches on their own.

These findings suggest a potential complementarity between the statistical approaches, rather than requiring that a researcher rely solely on one paradigm and not the other. Such methodological insights should be of interest to scholars over a range substantive applications, given that synthetic control methods have been used widely in recent years in influential research that covers a vast array of research questions: the influence of natural disasters on longterm livelihood (Cavallo et al., 2013), the effects of terrorism on local-level growth (Abadie and Gardeazabal, 2003), how political term limits shape decision making (Keele, Malhotra and McCubbins, 2013), the effect of border reforms on the rate of illegal immigration (Bohn, Lofstrom and Raphael, 2014), the role of political

connections for corporate valuations (Acemoglu et al., 2016), the effect of organized crime on longterm economic growth (Pinotti, 2015), how consumer boycotts influence international trade (Heilmann, 2015), the degree to which gun control laws reduce in-state homicide rates (Rudolph et al., 2015), how the imposition of compulsory voting laws shape democratic outcomes (Fowler, 2013), even how of the transfer of superstar athletes may shape sports leagues (Kleven, Landais and Saez, 2013). The list goes on. These diverse settings, while substantively unrelated, are unified by a researcher's urge to draw individual-level causal inferences when treatment is uncommonly assigned, when the set of possible control units does not closely resemble the exposed unit, and when data are collected in a time series. Such conditions, as discussed in (Abadie, Diamond and Hainmueller, 2010), tend to impede traditional inferential techniques, which has drawn well-deserved attention to flexible techniques such as the Abadie, Diamond and Hainmueller (2010, 2015) synthetic control estimator.

Results from observed data and simulation studies suggest the BSTS is an appropriate methodology for estimating individual-level causal effects in event-study type designs. When the synthetic control estimator tends to perform relatively poorly, the BSTS tends to performs relatively poorly as well, but it tends to nevertheless exhibit superior model fit the than the traditional synthetic control estimator on its own. The poor performance of the synthetic control estimator does not guarantee that the BSTS will also perform poorly, however. (To see this, consider the case in which an exposed unit's counterfactual series is perfectly negatively correlated with unexposed series.) The BSTS tends to be weakly improved when the two approaches are used in tandem, however, because the model's variable selection stage may accept or reject a potential covariate as a function of that variable's informativeness. Tasks for future research include an inspection into the performance and calibration of BSTS models for differing data types (e.g., discrete data) or how relaxations of the SUTVA assumption may be embedded into analysis. Replications of prior synthetic control designs (such as those outlined above) may also provide insight

as to whether modeling choices undesirably lead to biased causal estimates or differing substantive conclusions.

# Chapter 2

## Dimensions of Diplomacy: Understanding Private Information in U.S. Foreign Policy Using the WikiLeaks Cable Disclosure<sup>1</sup>

### 2.1 Introduction

*“[T]he nature of foreign negotiations requires caution; and their success might often depend on secrecy; and even, when brought to a conclusion, full disclosure of all the measures, demands, or eventual concessions which may have been proposed or contemplated would be extremely impolitic: for this might have a pernicious influence on future negotiations; or produce immediate inconveniences, perhaps danger and mischief, in relation to other Powers.”*

— George Washington, Speaking to the House of Representatives, 1796

State secrets and the private information possessed by leaders are nearly impossible to observe in practice (see Colaresi, 2014: for an overview). Yet such quantities—and the general difficulty of communicating that information in the form of a state’s capabilities and resolve—are at the core of much of the modern study of conflict in international relations (Frieden and Lake, 2005; Lake, 2010). In that paradigm, private information is a crucial ingredient of understanding why disputes may arise (see, e.g., Powell, 1999, 2002;

---

<sup>1</sup>Co-authored with Arthur Spirling. Erin Baggott, Amber Boydston, Andy Hall, Robert Schub, and Anne Sartori provided very helpful comments on an earlier draft. We thank audiences at the Harvard Political Economy Workshop, Columbia University, MIT, Princeton University, the New Directions in Text Analysis conference, and the Midwest Political Science Association meeting for feedback.



Tarar and Leventoglu, 2009) and escalate (e.g., Fearon, 1994a), how they might end (e.g., Goemens, 2000), their duration (e.g., Slantchev, 2004) and whether they might be avoided in the first place (see, e.g., Fey and Ramsay, 2011). This central place for information, secrecy, and beliefs also holds in theories of international relations that do not invoke the tenants of bargaining directly, including those that rely on ideational motivations (e.g., Wendt, 1999) or the pursuit of material resources and state security (e.g., Waltz, 1979), more broadly. Perhaps unsurprisingly given the importance of communication and information control, recent theoretical research has turned particularly to the use and practice of diplomacy in the international system (see, e.g., Sartori, 2002; Smith and Stam, 2004; Sartori, 2005; Kurizaki, 2007; Trager, 2010; Ramsay, 2011). There it joins a now well-established empirical literature that either explores the plausibility of the fundamental tenants of the rationalist approach (e.g., Fearon, 1994b; Partell and Palmer, 1999; Werner, 1999; Reed, 2003; Ramsay, 2008; Reiter, 2009; Potter and Baum, 2014), or assuming that those assumptions are correct, gives methodological advice on how to fit statistical models consistent with them (see, e.g., Signorino, 1999).

Despite clear progress made on creating and examining models that do a better job of describing the nature of crises as observed by international relations researchers, scholars still understand relatively little about the practice information transmission and protection in less extreme, non-crisis settings—events that constitute the majority of everyday international interactions. Accordingly, outside of some specific policy areas relating to conflict (on military operations, see, e.g., Keohane and Nye, 1977) we have a dearth of knowledge regarding the more general diplomatic behavior of leaders and bureaucrats around the world. For example, we know very little, in an empirical sense, of how ‘military capabilities’ are actually conceived of by official agents, how such information is withheld or protected relative to other international political issues, and how that conception affects what is promoted, concealed or communicated to foreign and domestic audiences. This situation is unsurprising, but unsatisfactory. It is unsurprising—

given limited resources, including researcher time—the literature has focused on crises and war, which have profound welfare consequences. Assuming one did want to study more general interactive practices, data limitations are prohibitive in any case: secret information is, by definition, closely guarded by states and even when their files are declassified and actors willing to give interviews, they do so in an obviously selective way (see, e.g., Shapiro and Siegel, 2010: for discussion). Furthermore, were such data possible to obtain, it is not in an obviously usable form. In particular, quantitative scholars have tended to hone their techniques for observational data in which each unit represents the incidence of a particular phenomenon or event of interest (e.g., Ghosn, Palmer and Bremer, 2004), whereas information pertaining to diplomacy is mostly in terms of documents (primarily cables) sent between embassies and bureaucratic departments and ministries. In such a world, what constitutes an individual observation is quite unclear.

This state of affairs is unsatisfactory for more obvious reasons: put very crudely, as a discipline, we do not know how secrecy actually behaves in contemporary diplomacy, despite the fact that it is a vital part of countless relations theories. Our ignorance in international relations regarding such a core component of scholarly theories may be compared with a much more favorable situation in related areas of political science which have new and fine-grained data to assess the plausibility of their models, and to expand their understanding of the processes therein.<sup>2</sup>

This paper introduces new data and methods to get precisely at these fundamental issues for United States foreign policy: that is, we characterize diplomacy in terms of what is kept secret, and provide explanations as to why. Our data are 163,958 cables dealing with the period between 2005 and 2010—an era in which coverage is relatively dense, and during which the United States had several ongoing military operations in the Middle East. The cables—which are essentially secure emails sent between the

---

<sup>2</sup>As an example, consider the now voluminous literature on psychology in international relations (starting at least with Jervis, 1976), in which scholars have used new measures of biological responses to assess the effects of emotion on political opinions (e.g., Renshon, Lee and Tingley, 2015).

Department of State and its embassies and missions in hundreds of cities around the world—include large numbers of documents never intended for public consumption. As such, they include thousands of (officially) secret and confidential missives, and thus allow researchers extremely rare access to the world of private information in IR unfettered by state censorship. This paper’s primary theoretical contribution is to argue that private information is as much about ‘procedure’ as it is about ‘substance’. Put differently, though the United States certainly does not wish other state actors to discover certain facts about its material capabilities (especially regarding matters of ‘high politics’ in the sense of Keohane and Nye, 1977), it may also wish to obfuscate the way that it allocates intelligence gathering resources. That is, the United States seeks to hide the information it gathers, from whom it garners it, and what it chooses to disseminate. We call this latter dimension of diplomacy ‘procedural’ and contend that it refers not to actions regarding specific objects (such as arms, services, plans or strategies) but rather to a method of behavior in general, regardless of political issue areas. In keeping with a pre-existing literature on diplomacy (particularly Sartori, 2002), we provide corroborative evidence for the notion that developing and maintaining a reputation for confidentiality matters to diplomats and their staffs. We go beyond current accounts, however, in demonstrating that patterns of information protection hold somewhat independently of the policy areas discussed.

While the purpose of this paper is not to ‘test the assumptions’ of strategic models *per se*, our hope is that our inspection will aid researchers interested in the empirical implications of such theoretical models. In particular, studying the way that information is shielded from global public view on an everyday basis provides a resource for those interested in ‘audience costs’ (e.g., Fearon, 1994a; Weeks, 2008)—which apply to crises bargaining situations, specifically. Furthermore, since we show that both procedure and substance matters for secrecy, we believe our empirical efforts provide an impetus for theory development in contemporary international relations research. This is in part

because our sample is much broader in substantive terms than has been available for previous studies, for which we have theoretical work already.<sup>3</sup> But it is also partly because the cables we use have ‘official’ designations in terms of their classification status and the justification for that status. That is, our work relies on the officially structured indexing of cables and their topical designations, for which state actors have made conscious decisions. Aside from data cleaning, we do relatively little to restructure the standardized form of the metadata observed on each diplomatic cable, as such a procedure could induce error in the variables we chose to measure. As a result, our inferences are less dependent on somewhat idiosyncratic or arbitrary research rules, and thus sharper than previously possible; we believe this makes them especially ripe for theorizing (for further testing thereafter), as our topical policy codings are *identical* to the internal indexing structure used for information management at the State Department.

In undertaking the study, our paper contributes methodologically and suggests new ways of working with ‘texts-as-data’ (Grimmer and Stewart, 2013*b*). In particular, we use machine learning techniques—such as random forests (Breiman, 2001) and the ‘lasso’ (Tibshirani, 1994)—in tandem with matched sampling designs to identify how ‘important’ terms discriminate between restricted and unrestricted documents. We provide novel ways of comparing texts, based on matching on the metadata of each document, such that political scientists may think sensibly about the (marginal) influence of secrecy on a document’s content.

## 2.2 The Study of Private Information Disclosure

To get much of their purchase on the world, the conclusions drawn from formal theories of international dynamics often hinge on differences between private and public information,

---

<sup>3</sup>Our efforts are somewhat similar in spirit to those of Weidmann (2016) who uses previously restricted military data to study shortcomings in the reporting of violence in an effort to improve empirical practice in the area.

and the degree of information overlap or ‘common knowledge’ shared by actors in a non-cooperative environment (see Powell, 2002: for an overview). Somewhat understandably, models in this tradition rarely define exactly what constitutes the manifestation of information of either type. Primarily this is a data issue: understanding what states know, what they do not, and what they are keeping secret cannot be determined deductively insofar as official refusal to answer queries about particular issues does not allow scholars to become much more informed about the true state of the world. While it is correct that documents pertaining to international decision making are routinely declassified, this process tends to be slow, somewhat haphazard, and obviously case selective (see Allen and Connelly, 2015: for an overview of U.S. protocols).

Given data limitations for recent historical periods, scholars interested in examining the plausibility of theories that rely on information—for example, those that utilize ‘audience costs’—tend to pursue one of three avenues. First, they use observational (or survey) data in a regression context (e.g., Weeks, 2008). This has the obvious benefit of being straightforwardly replicable, but arguably lacks the kind of internal validity that would be convincing for skeptics (see, e.g., Trachtenberg, 2012: for recent discussion). Second, they enter the archives of governments and produce historically rich case studies (e.g., Schultz, 2001; Snyder and Borghard, 2011), which are necessarily limited to specific times and places. Third, they undertake field or survey experiments on political actors or on less representative samples (e.g., Tomz, 2007). In all cases, if researchers had broader access to cases or incidences of censored versus public materials, they might be able to draw sharper conclusions—better in terms of both internal and external validity—than current approaches allow.

Whatever the research approach, a crucial insight of the audience cost literature is that increasing the publicity of some political matters may be useful, particularly if the issue at hand is one of signaling a commitment credibly—and we might expect variation in a nation’s willingness to publicize issues given the strategic importance of the issues at

hand (e.g., Fearon, 1994a). Of course, strategically making information public is helpful beyond the signaling case: citizens sometimes need to be warned to avoid possible threats, just as investors may be influenced by the announcement of shifting taxes or interest rates. Along these lines it is not hard to think of ‘substantive’ facts, especially connected to capabilities, states may want to conceal—for example, where nations locate nuclear warheads, how many they have, or how easy they are to launch. It would directly hurt a state were any of this information known, since it compromises defense plans, decreases combat effectiveness, and broadly provides potential enemies with additional bargaining power.

Inasmuch as nations have incentives to disclose some information but keep other information private, one can imagine scenarios in which governments prefer to conceal what they are ‘trying to know’ or how they acquire, discuss, or share information independent of any substantive issue at hand. In brief, there may be incentives to practice discretion, on average, independent of a particular policy issue being discussed diplomatically. Suppose, for example, the United States had a rule to publicize its diplomatic communications on some policy areas but not on others; competing nations could reasonably update their beliefs on what is likely being negotiated or communicated in private given the absence of that policy in being revealed in public summaries. In such a world, one can see how a practice of selective disclosure could in fact advantage international competitors relative to a more general norm of secrecy and discretion. In an extension of this logic, nations may have incentives to conceal information concerning the meetings of leaders or public officials irrespective of the topics or issues discussed.<sup>4</sup>

---

<sup>4</sup>Taking this further, suppose it was common knowledge that leaders scheduled a meeting on a certain day to discuss an unknown issue; if a rule required that the topic of the meeting be made public if it concerned certain political issues (e.g., the meeting’s topic is revealed if about environmental politics, but kept private if about security concerns), then the absence of a public summary of the meeting would nevertheless reveal information to the public about the likely issues raised in private—a fact that could make a rule of selective disclosure, at least on the margins, strategically self-defeating. It would be more difficult for the public to update its beliefs on the likely content of a meeting if no such (disclosure) rule were to exist, by contrast.

Although the discussion above concerns the practice of disclosure in international diplomacy, the intuition that selective (i.e., issue-specific) information disclosure may be welfare reducing is at the core of a much broader class of political phenomena. There is a vast microeconomic literature on the public dissemination of private information in macroeconomic environments with complementarities—an area of economic research which tends to focus on the conditions under which information disclosure may be welfare improving or influence a system’s volatility. Common settings of such studies concern central bank communications and global financial systems, where decision makers may wish on the one hand to disseminate some information to public audiences but avoid being so transparent as to induce crises or speculation on the other.<sup>5</sup>

### 2.2.1 Implications and Expectations

In international relations, communications between diplomats and state officials may be in service of multiple ends. Diplomats may wish to signal their intentions about U.S. foreign policy to other leaders, share sensitive information to trusted confederates, and—in perhaps an ideal case—update the quality of their private information about political issues that might otherwise be difficult to observe through other means. With that said, there is no perfect way to observe all private information available to diplomatic actors through records of communications and their associated handling statuses. Even if

---

<sup>5</sup>In the determination of monetary policy, bankers face decision problems that resemble those of international diplomats, although this parallel is not a focus of the literature. When making a public statement, bureaucrats may be concerned both with the precision of a public signal (i.e., how closely a message will map to real outcomes) in addition to the degree of its publicity (i.e., how many individuals observe a signal). Cornand and Heinemann (2008) provides a careful discussion as to how precision and publicity may shape the decision to disseminate private information. The authors argue, being in step with much of the theoretical work related to this problem, “The optimal degree of publicity depends on the precision of announcements” (718)—more specifically, that if the precision of a public signal is not guaranteed to be sufficiently high (e.g., if the event or issue described in a communication is not sufficiently likely to occur in reality), it may be dominant to avoid private information dissemination altogether. Arguments more favorable to how transparency may be welfare improving are in Cukierman (2001) and Angeletos and Pavan (2004), under the requirement that the quality of public signals is high. The model presented in Woodford (2005) provides a sufficient condition for when bureaucratic transparency should lead to a welfare reduction in expectation.

our data contained records of all top-secret communications between relevant officials, our sample would still be unable to sensibly measure all officials' private information *per se*.

What can be approximated, however, is the degree to which particular policy issues and content-based features map to higher levels of political protection within our sample. Common sense dictates the relationship between political issues and handling status ought to be related to political actors' preferences over the sensitivity of a political issue area, or more broadly, the incentives to shift information disclosure to the public on a specific political issue. If we are to observe that some political issues are systemically more predictive of document secrecy than others, *a fortiori* this provides evidence of diplomatic preferences over the sensitivity of political issues in the international system.

Formal theories of strategic information disclosure suggest diplomats would have incentives to withhold information as a function of the presumed cooperativeness of a decision environment. All else equal, as incentives between nations to coordinate on policy issues decrease (e.g., on national security discussions or information on nations' relative capabilities), we would expect communications on such issues to be more protected in our sample on average than policy areas with more of a cooperative or 'common-pool' character (such as environmental concerns). In terms of their rank ordering, a natural prediction concerning the 'substantive' secrecy of our sample would suggest a monotonic decrease in the in-sample estimates of cable classification as one moves from traditionally 'non-cooperative' games on one end to 'cooperative' issue areas on the other.

To assess the plausibility procedural dynamics as a driver of information restriction, our aim is to extract features of language that are predictable of cable secrecy after adjusting for the topical focus of diplomatic communications. If formal theories on information dissemination focus on minimizing enemies' abilities to anticipate foreign decisions, one might expect particular textual features of diplomatic cables to be predictive of secrecy conditional on subject matter. In particular, if meetings between diplomats,



ambassadors, or leading political officials serve an information-gathering purpose, one might expect references to such individuals, all else equal, to be associated with a document's probability of restriction.

## 2.3 Data

In 2010, a data breach at the U.S. State Department led to the public release of 251,237 diplomatic secure messages sent by the U.S. State Department to U.S. embassies and missions. The date range for the original data is from 1966 to 2010, and in Figure 2.1 we plot the total number of cables per month from that time period. In this study, we focus on all cables written and sent between January 1, 2005 and the end of the data, through February of 2010. We do this for two reasons: first, coverage prior to the year 2000 is somewhat sparse and inconsistent. Second, because we were concerned about changes to security procedure (e.g., requirements to copy-in embassies and missions on particular types of messages) after the terrorist attacks of September 11, 2001. This leaves 163,958 documents from which to draw inferences.<sup>6</sup>

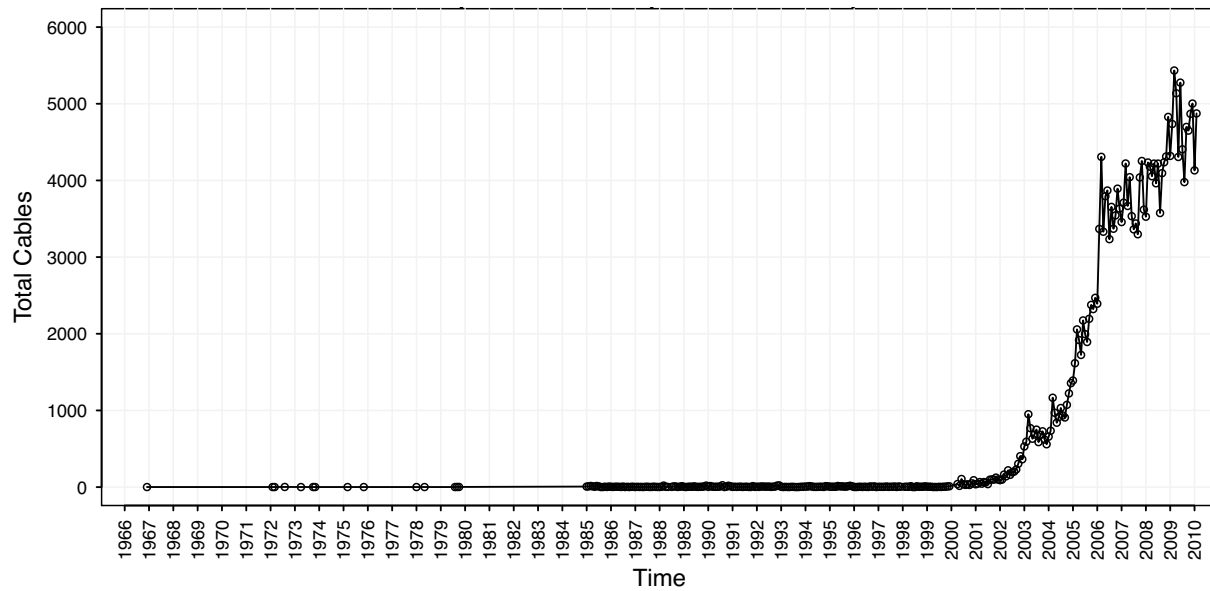
According to official rules, cables may be classified into one of three categories, depending on the degree of damage to national security that "the unauthorized disclosure of which reasonably could be expected to cause."<sup>7</sup> Furthermore, any classified document must pertain to at least one of a series of topics which *inter alia* include military plans, intelligence, foreign relations of the United States, nuclear programs, weapons of mass destruction and vulnerabilities in national security. In descending order of the purported balefulness of unauthorized release, these categories are 'Top Secret', 'Secret' and 'Confidential'. If a cable does not meet the criteria for such restricted access, it is deemed 'Unclassified'. In our particular data, we have the following distribution: zero

---

<sup>6</sup>Gill and Spirling (2015) provides a discussion of the "representativeness" of the this sample's disclosure, but no detail as to the actual content of the diplomatic cables.

<sup>7</sup>As described in Executive Order 13526, 2009.

Figure 2.1: Number of cables per month in Sample, 1966–2010



Notes: While the cables disclosed span decades, the post-2005 period has much more dense coverage, and is therefore the focus of our inquiry.

Top Secret, 10,195 Secret cables, 87,270 Confidential cables, and 66,493 Unclassified. There are, in addition, some extra classifications that appear less frequently in the data, such as ‘Confidential and Not For Foreign Distribution’, ‘Unclassified for official use only’, and ‘Secret and Not for Foreign Distribution’; we ignore these categories for our current efforts.

For our purposes below, we divide the categories into ‘restricted’ (R), which includes Secret and Confidential communications, and ‘unrestricted’ (U), which includes the unclassified documents only. The central idea here is to separate documents into more ‘private’ and more ‘public’ information, respectively. This measure is somewhat coarse, but given that theories in International Relations use similarly binary demarcations we think this is reasonable. To be clear, the fact that a cable is unrestricted does not imply it is automatically made public: it is still a government document rather than a press release. But unclassified documents—so long as they are not ‘For Official Use Only’—do

make their way into the public domain, and are eligible for release under Freedom of Information Act requests. Put otherwise, our unrestricted case covers documents that the public (anyone without specific security clearances) could access; our restricted cables are those that are not released or releasable to the public.

Any given document may be given 'tags' assigned by its authors, with guidelines for this process provided by the State Department.<sup>8</sup> From our perspective, these tags contribute meta-data that communicates the topic of the content therein, and are assigned to each document following its completion. After cables are written and subject tags are assigned, each cable is given its overall classification status.<sup>9</sup> Examples of subject tags in our data include 'ADCO' which refers to 'Diplomatic Courier Operations', 'PTER' which refers to 'Terrorists and Terrorism', 'SMIG' which pertains to 'Migration' and so on. There are a total of 97 tags in our data, though their use varies widely in relative frequency terms. The full list can be seen in B.4. The variety in tag number per document can be seen in Figure 2.2; inspection suggests that the modal number of tags is two or three, though there are 14,451 unique combinations of subject tags (ignoring each cable's location of origin) that appear in our post-2005 sample at least once.

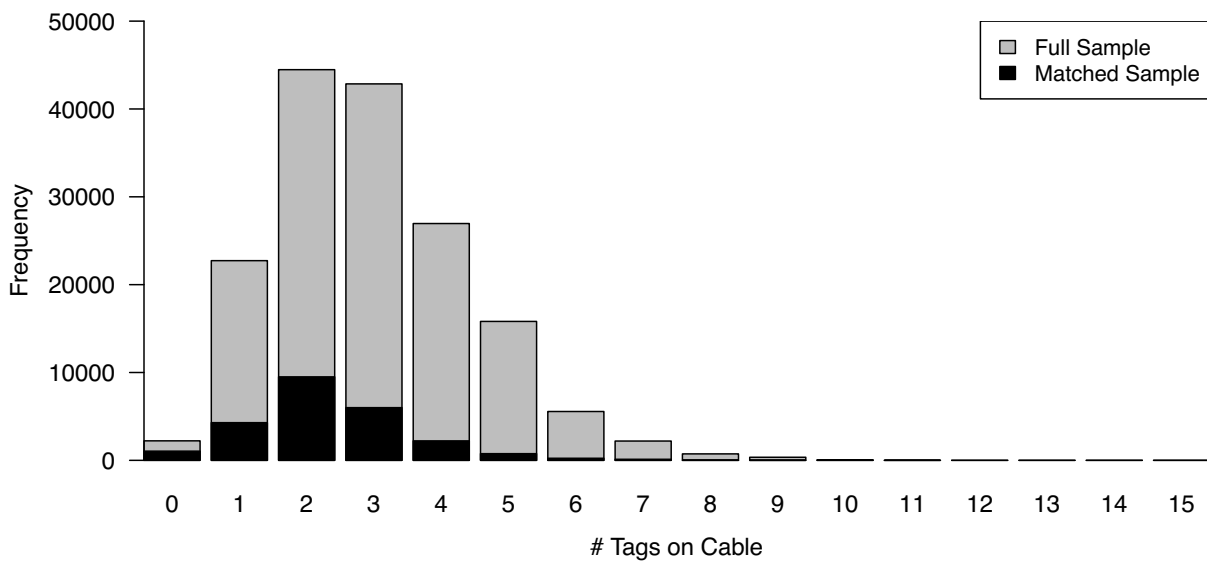
In Figure 2.3 we report the structure of the data in terms of the way that tags co-occur across cables. Areas of darkness in that plot are places where tags coincide. Our main observation is that tags in section 'P' (which denote 'Political' issues) and, to a lesser extent, tags in section 'E' (denoting 'Economic' matters) tend to coexist heavily with other subject indicators, suggesting that these issues play an important organizing role in the U.S. diplomatic service. Machine readable versions of the documents themselves are

---

<sup>8</sup>These are literally geopolitical 'TAGS', an acronym for 'Traffic Analysis by Geography and Subject', implemented for diplomatic communication in its modern form by an executive order (number 11652) in June 1974. Their justification was to "[p]ermit more rapid and discriminating distribution of messages", and to "[p]rovide statistics to both offices and posts on what is being communicated in the Department-field system"; they were to "[s]erve as headings for clustering the terms used by professional indexers to identify the content of substantive messages."

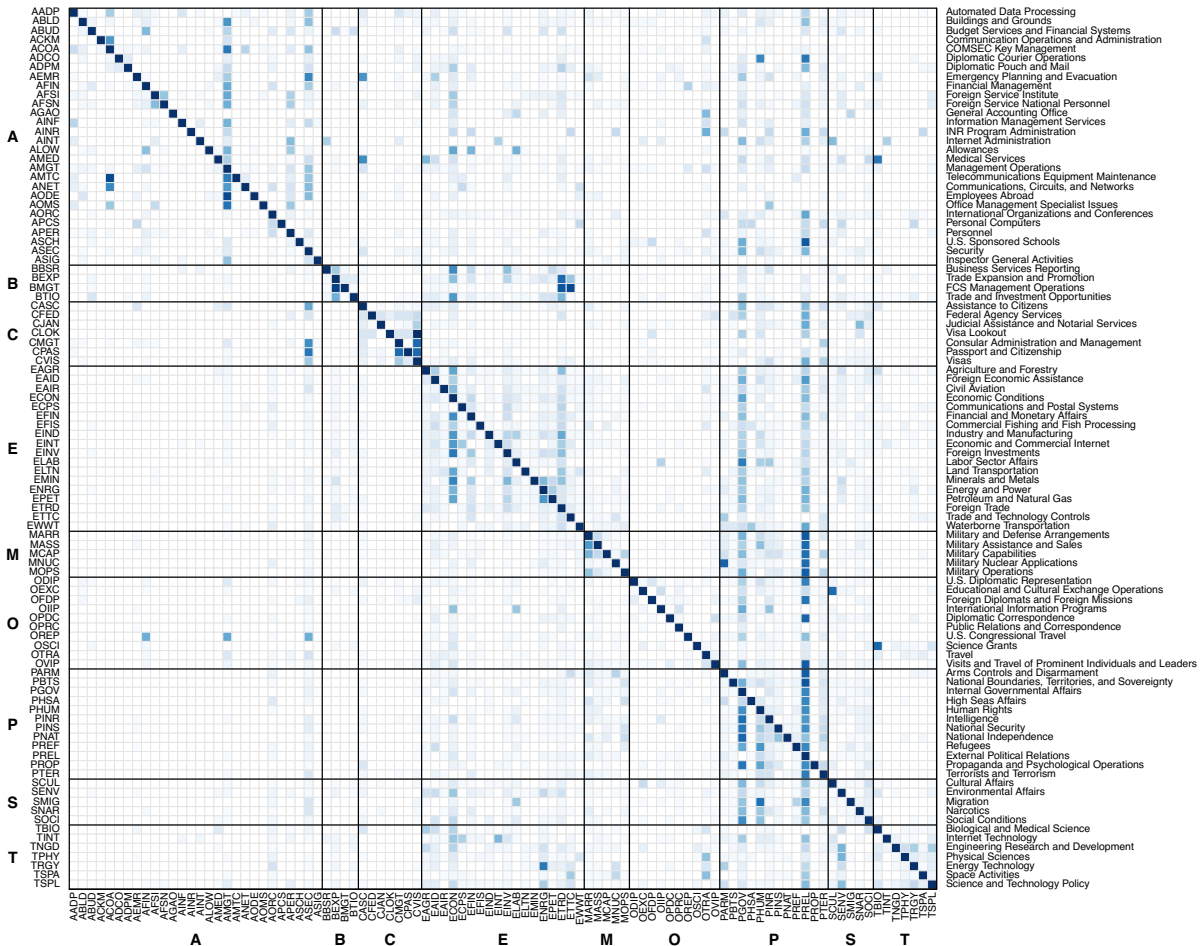
<sup>9</sup>Additional details on the origins and formal procedures of cable creation can be found in B.2.

Figure 2.2: Distribution of tag counts in sample



Notes: The modal cable in the data has two subject tags on it, although it is not uncommon to see cables with four or five subject codings. Details of the matched and unmatched study samples are discussed in Section 2.4.2.

**Figure 2.3: Conditional probability of U.S. State Department subject TAG co-occurrences in the post-2005 sample (n = 163,958)**



*Notes:* This figure shows the (empirical) conditional probability of U.S. State Department subject TAG co-occurrences in the post-2005 sample (n = 163,958). Subject tags are presented in alphabetical order with their official U.S. State Department meanings listed in the righthand column. Each cell in the figure represents the conditional probability that a *column subject* will be tagged given that a *row subject* has already been tagged. Darker shaded cells indicate a higher conditional probability.

available at various websites for download, though some pre-processing is then required prior to any analysis. In particular, the tag information must be captured and removed, and some other cleaning performed. Much of what follows involves operations on the ‘document-term matrix’ (DTM) of the texts, which was ‘stemmed’ (meaning that words were pruned back to their ‘roots’ where possible, using the Porter (1980) algorithm), ‘stopped’ (meaning that function words which are thought to contain little discriminating information were removed), and subject to a ‘sparsity’ condition of 99-percent (i.e., only words that occur in more than one-percent of all documents but in no more than 99-percent of all documents are included). Such data cleaning is common in text analysis in the social sciences (e.g., Grimmer and Stewart, 2013b). The resulting DTM for analysis is matrix with dimensions  $163,958 \times 3,755$ .

## 2.4 Methods

Our claim above is that the secrecy endemic to diplomacy comes in at least two separable varieties: substantive secrecy—the notion that certain information about a policy area is to be kept confidential because it would be *per se* damaging to security were it released—and procedural secrecy, which is concerned with the notion that secrecy protects foreign or domestic agents from outside consequences of their actions. To assess the evidence for these separate ideas, some care is required in terms of methods. Here we explain our approaches.

### 2.4.1 Substantive Secrecy

We first examine the question of *substantive secrecy*—i.e., how a State Department topic or substance of a diplomatic communication, all else equal, influences its probability of restriction. The objective is to quantify both the magnitude and direction of how the presence of official U.S. State Department communication subject tags influence cable

secrecy. In suit, we regress each cable's observed restriction status on its subject tags and location of origin. This fixed-effects least squares equation can be written as follows:

$$R_i = \alpha + \sum_t \beta_t Tag_{it} + \gamma_j + \varepsilon_{ij} \quad (2.1)$$

where  $R_i$  is a dummy variable for cable  $i$  that takes the value of 1 if the cable is restricted and 0 if unrestricted,  $Tag_{it}$  is a subject tag dummy variable for cable  $i$  for each tag  $t$ ,  $\gamma_j$  is the fixed effect for embassy  $j$ , while  $\alpha$  and  $\varepsilon_{ij}$  are the constant and error terms, respectively. Given that each covariate in this regression is binary, each regression coefficient  $\hat{\beta}_t$  is a sample estimate of the difference between two conditional expectations: the conditional probability a document will be restricted given the presence of a subject tag minus the conditional probability of restriction without that subject tag present.<sup>10</sup>

## 2.4.2 Procedural Secrecy

Recall that procedural secrecy concerns the diplomatic norms of confidentiality in meetings. If it exists as a quantity that can be identified in our data, then it should emerge as a key discriminator between restricted and unrestricted cables. However, if there is indeed a subject tag imbalance between restricted and unrestricted cables (as suggested above), this implies that a simple comparison of word frequencies between restricted and unrestricted documents is unlikely to isolate how text varies on the margins a function of secrecy status alone, since observed differences are likely to arise directly from ex ante differences in subject matter.

Thus, the question we ask in this section is: having adjusted for cable subject matter (given an observed sequence of subject tags on a document) and locations of origin, all else equal, can restricted diplomatic communications be distinguished from unrestricted

---

<sup>10</sup>Although the outcome of interest is binary, a linear model is appropriate when the conditional expectation function (CEF) of each regressor with respect to the outcome exhibits is linear (see, e.g., Angrist and Pischke, 2009: Chapter 3). OLS suitably estimates whereby sample average effect of restriction on each subject tag in the context of our data, as each regression coefficient represents a conditional mean.

communications? This question may be thought of as estimating the *marginal effect* of secrecy on the content of a restricted communication. More precisely, given two documents indexed with identical subject tags and originating from the same source, are there specific textual features that systematically distinguish restricted cables from unrestricted cables? If such textual features exist, is there anything substantively unifying about these features? In particular, does whatever differentiates these communications be considered ‘procedural’ in nature?

### **Exact Matching on Subject Tags and Origin**

To assess whether secrecy, on the margin, is associated with differences in document composition, we restrict our sample to *exactly matched* subsets of cables within each embassy in our sample. More precisely, for each embassy (i.e., each cable’s location of origin), we implement the algorithm outlined in Figure B.26 in B.8 to construct datasets of cable pairs that are exactly matched on official U.S. State Department subject tags and their embassies of origin, but differ on their restriction level. The objective of this matching procedure is to restrict the full sample such that there is perfect subject overlap on cables in our study. As a result of the matching procedure, within each embassy, for each restricted cable there will exist an unrestricted cable that has an identical subject tag pattern. We rely only on the State Department’s official subject tags for this procedure. If two or more unrestricted matches are found for a single restricted cable, we select the match that is written most closely in time to the restricted cable’s date of authorship. For the results presented in this study, matching is performed without replacement. Datasets are stored and analyzed both at the embassy level and in pooled analyses.

Since we wish to make inferences about textual differences between restricted and unrestricted cables on the margin—i.e., once cable subject tags have been accounted for—the within-embassy matched sampling design has clear appeal and allows for a meaningful examination of procedural secrecy. Adjusting the sample directly for



differences in subject matter and controlling for embassy-level effects, the design allows us to isolate differences in textual composition that are likely to arise from a document's handling status alone. Intuitively, the aim of our exactly-matched sample is to "control" for substantive differences in cables that may be present in the unmatched sample—differences that may arise from hypothetical variation in reporting rules, document disclosure standards, authorship style, or political priorities at the embassy level. If systematic textual differences remain between restricted and unrestricted cables after subject and location have been accounted for, these differences are likely to arise from residual rules that are separate from subject-specific handling rules.

The formal appeal of exact matching is that it is nonparametric and approximates the act of "blocking" in randomized experiments (Cox, 1958; Imai, King and Stuart, 2008).<sup>11</sup> Exact matching is often untenable in applied research, however, since in many cases the sampling procedure can dramatically reduce a researcher's final sample size, and the procedure tends to rely on initially large sample sizes. Unsurprisingly, this was a concern for our modeling attempts, along with the possible danger that many documents dealing with sensitive substantive areas would be jettisoned from the final analysis because no match could be found for them. Further, we were concerned that certain 'important' embassies would be, relative to the original dataset, heavily under-represented.

Neither of these concerns appear to be true of the study sample. In B.6 we report the reduction in subject tag imbalance of the exactly matched sample, in addition to information on which subject tags remain present. In the exactly-matched sample, we see both embassy-level and aggregate level subject imbalances have been eliminated.<sup>12</sup> In B.7

---

<sup>11</sup>In an exact covariate matching procedure, if the appropriate set of conditioning measures has been identified, the unobserved functional relation between between covariates and the assignment to treatment is ignorable due to *perfect balance* on conditioning variables. Under general conditions, exact matching procedures are both *equal percent bias reducing* (Rubin, 1976) and *monotone imbalance bounding* (Iacus, King and Porro, 2011). These traits are not generally true for most distance-based or model-based (parametric adjustment) matching methods, which has led several scholars to conclude that exact matching is close to an "ideal" matching procedure in observational settings (e.g., Stuart, 2010; Imai, King and Stuart, 2008).

<sup>12</sup>Exact matching will allow us to inspect textual differences akin to *treatment effects on the treated*.

we report the embassies, and their relative prevalence, in our matched data. Importantly, we note that ‘larger’ embassies—including the U.S. State Department itself—are most represented; in particular, Ankara, Baghdad, Paris, Cairo and Moscow (all centers of activity in the original data) appear at higher rates in the matched sample. Taken alongside the successful the reduction in subject imbalance, this presents strong evidence that the matched sampling procedure does not leave the general patterns of the whole sample too far behind, and is due to the fact that there are sufficiently high within-embassy subject tag correlations. The diplomatic locations contained in the matched sample are represented in a manner proportionate to their overall representativeness in the post-2005 sample.

### **Supervised Learning and Penalized Regression**

For each of the matched samples described in Section 2.4.2, we implement a set of supervised learning models to identify which words are most *important* to (i.e., predictive of) cable secrecy. The matrix of words used in this classification setting is taken from the full post-2005 document-term matrix described before, but now only includes rows that satisfy the within-embassy, exactly-matched sampling design. On the ‘left hand side’ we have the (binary) restriction status of a given document which we intend to predict with the words within that document. Quantitatively, we observe how within-sample classification error rates vary as a function of which words are included in the model; qualitatively, we wish to make statements about how a document’s restriction status would likely change if particular words within these documents were to vary. Two supervised learning methods are applied to these data: the “random forest” (hereafter RF) algorithm

---

Treatment effects on the treated are not the same as (unconditional) average effects, nor are they the average treatment effect in the sample. More precisely, they concern how potential outcomes would differ for a set of treated units in the sample if they were instead to become untreated. In the present study, therefore, the design allows us to estimate answers questions like the following: If a set of treated restricted documents like those in our sample were instead to become unrestricted, on what textual dimensions would we expect those collections of documents to vary?

(Breiman, 2001), and the “lasso” (Tibshirani, 1994). Results from both procedures are used alongside the topic model estimates described below to make statements both at the world-level and topic-level about how secrecy, on the margin, influences the content of diplomatic communications. More details on the RF and lasso procedures can be found in B.3.

With both RF and the lasso, we obtain embassy-level estimates of word-level dependencies to document restriction. In the context of RF, each exactly-matched dataset for embassy  $j$  has a corresponding vector of *word importances*, where importance is defined as an estimate of each variable’s in-sample average marginal error reduction. In the context of the lasso, each embassy has a corresponding vector of penalized partial regression coefficients. For both the RF and the lasso procedures, we refer to this collection of embassy importance vectors as the *embassy importance matrix*. Each row in this matrix represents a given embassy, and each column is a measure of a word’s relative importance to prediction accuracy in the embassy’s matched sample. Each cell entry is then the RF importance measure for that term in that embassy. To obtain sample-average estimates of word-level importances to prediction, we weight the results of each embassy-level importance vector by its relative share of all cables in the exactly matched sample. The prevalence of any given embassy in the matched sample, therefore, proportionately weights the importance terms associated with that embassy (thus, for example, we will up-weight the importance terms associated with the State Department itself and other embassies near the top of Figure B.25, as in B.7). Using the sample-weighted results of the RF within-embassy, exactly-matched classification procedure, we then took the top 30 of these terms (recall that they are all positively signed, regardless of their actual signed effect on classification), and recorded their corresponding coefficients from the lasso regressions at the embassy level. The lasso regression coefficients are similarly weighted as sample averages in proportion to each embassy’s representation in the matched sample.

## Supplementary Analysis: Topics

Some supplementary analyses are performed to address what differentiates more restricted documents from less restricted documents on the margins. In particular, we *topic model* our sample of  $n = 163,958$  cables, using the most common probabilistic topic model in contemporary text analysis research, Latent Dirichlet Allocation, henceforth referred to as LDA (Blei et al, 2003). Information on our topic modeling procedure is outlined in B.9. Results of the topic modeling procedure are used as an illustrative aid to categorize the words we find to be predictive of document restriction.

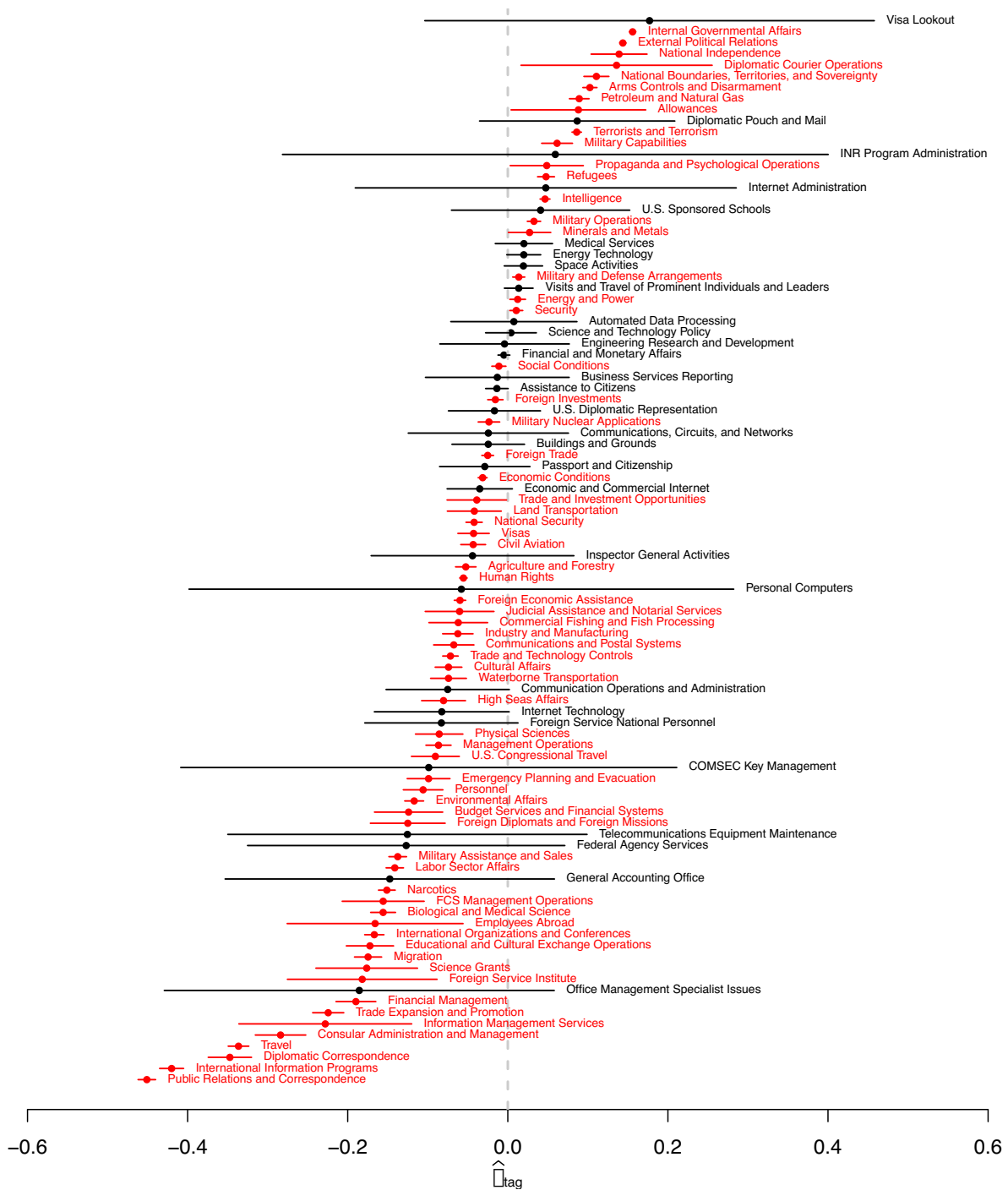
## 2.5 Results

We first interpret our tag regressions in terms of the nature of the substantive secrecy they reveal, before considering the evidence for our procedural secrecy hypothesis above.

### 2.5.1 Substantive Secrecy: High vs. Low Politics

Recall that testing for substantive secrecy boils down to testing whether or not the probability a diplomatic cable is withheld from the public is measurably predicted by the *subject* of the cable communication, adjusting for the cable's location of origin and other factors. Figure 2.4 presents this analysis, where each point corresponds to an estimate of the sample average effect of a subject TAG on the probability of the cable's restriction. Around each estimate is the 95-percent confidence interval. In terms of coefficient direction, note that the broken line in the center of the plot denotes a point estimate of zero 'effect': tags to the right of this line are generally associated with restricted documents (on average); the presence of tags to the left, generally predict an unrestricted status for the cables. Tags highlighted in red indicate coefficients that are statistically differentiable from zero. Our first observation is that there are a large number of statistically significant predictors: almost every subject matter tag is associated with increasing or decreasing the probability

**Figure 2.4: Substantive content as a predictor of secrecy status in Full Sample (Substantive Secrecy Analysis)**



Notes: Highlighted estimates are statistically distinguishable from zero after multiple comparisons correction (Holm, 1979). 95% CIs around each estimate.

that a particular cable is restricted. Second, we note that the direction of the effects are somewhat in line with our priors. Thus we see that cables concerning “Terrorists and Terrorism”, “Military Capabilities”, “Intelligence,” and “National Independence,” for example, are more likely to be kept private than cables concerning “Migration” “Narcotics,” “Personnel,” or “Environmental Affairs.” In particular, we see that dispatches dealing with ‘core’ state secrets, especially pertaining to information, capabilities and threats are restricted. We note that such subject matter accords with notions of ‘high politics’—specifically, state security and survival—as described by Keohane and Nye (1977). On the other hand, cables that discuss more ‘public good’ orientated matters—wherein we can imagine that sharing information may not be damaging, and may in fact be optimal—tend to be unrestricted. In this latter category are tags that seem to require or be synonymous with publicity and the dissemination of information: “International Information Programs”, “Public Relations and Correspondence”, “International Organizations and Conferences”, “Educational and Cultural Exchange Operations” and so on. With respect to the work of Keohane and Nye (1977), we might see such matters as ‘low politics’: issues of more domestic or economic concern.

The fact that cable substance drives at least some part of diplomatic secrecy should not come as a surprise to theorists of rational diplomacy. As noted above, most contemporary theoretical treatments of crisis diplomacy concern agents’ incentives to misrepresent their resolve, capabilities, or information in bargaining settings: our results here suggest the United States. acts in a way compatible with that logic.

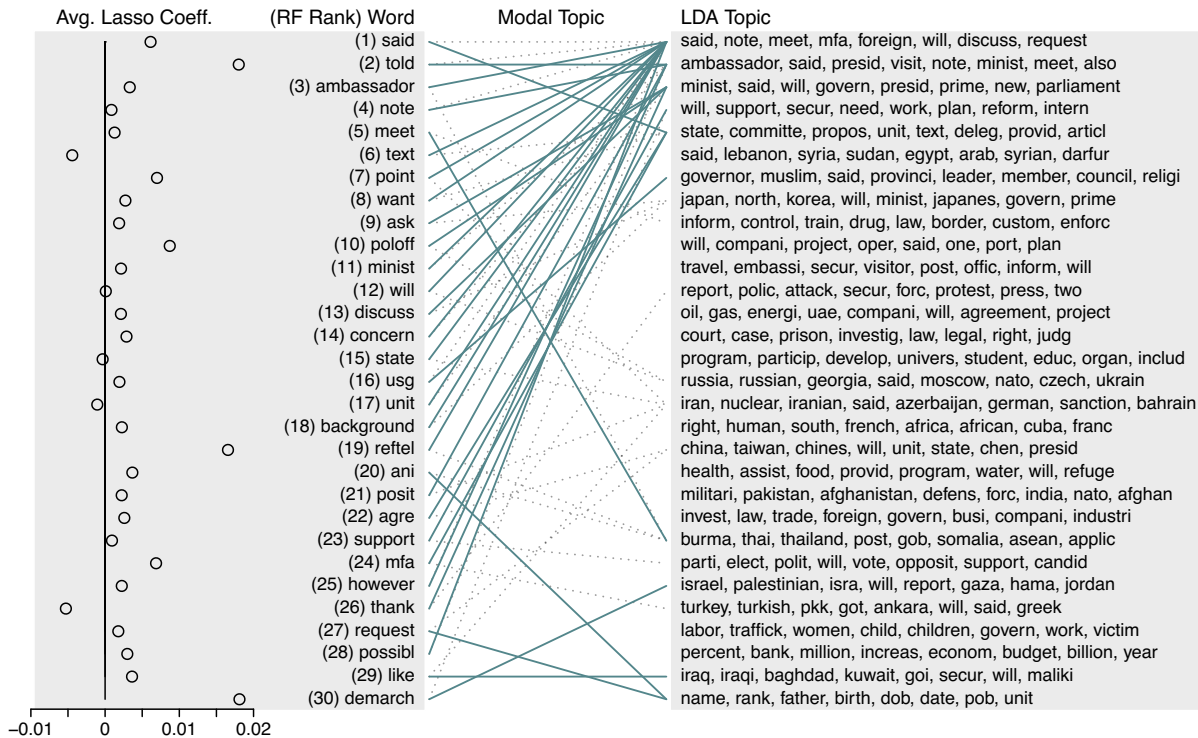
## 2.5.2 Matched Sample Results: Procedural Secrecy

In terms of procedural secrecy, an overview of our main results may be found in Figure 2.5. Recall that we used the RF algorithm to identify the thirty ‘most important’ tokens in terms of their ability to discriminate between the unrestricted and restricted cables status of a document. In the second column of the plot, these are clearly seen and include

words such as ‘said’, ‘told’, ‘ambassador’, ‘want’, ‘note’, ‘meet’, ‘want’, ‘ask’, ‘discuss’, ‘concern’, ‘state’, ‘agre[e]’ ‘support’, ‘however’, ‘thank’, ‘request’, ‘possibl[e]’, ‘like’ and so on. Our immediate observation is that in stark contrast to our tag regressions, these words do not connote substantive state secrets *per se*; rather, they refer to the holding of meetings and the general protocols of diplomatic exchange with foreign nationals. Related to this idea, note the presence of terms such as ‘poloff’ (the Embassy’s Political Officer), ‘usg’ (United States Government) and ‘minist’ (minister): actors who we expect to be involved in daily embassy interactions. On the left of the figure, we report the lasso (point) estimate associated with the terms. When these points are to the right of the vertical line, the use of that word (on average) increases the probability that a document is restricted; when the points are to the left, this suggests that the word is associated (on average) with a decrease in probability that a document is restricted. Examining this part of our results, we note that terms such as ‘said’ and ‘told’, ‘request’, ‘like’ are used disproportionately more in restricted cables. To us, this is evidence that once one controls for substantive area, secrecy is mostly about keeping meetings private and confidential, regardless of whether anything intrinsically ‘secret’ is being discussed.

To evaluate this intuition, we recorded the modal topic—i.e., for each word, the topic with the highest posterior probability from the topic model described earlier—in which our most influential words appeared. If we are correct that secrecy is partly about a norm of discretion rather than content, we would expect to see most of the terms mapping to a single (or perhaps a few) ‘administrative’ topic(s), rather than topics pertaining to matters of substantive import. On the right-hand side of the plot, we see this is almost entirely the case. There, the solid lines lead from each word to the topic it most likely belongs; the dashed lines are from each word to second most likely topic. We see first that with a few exceptions, all of the words ‘belong’ to the first, second, or third topics. Inspecting those more closely, we note that those topics generally consist of administrative nouns and verbs, rather than subjects of interest: thus, we find “said” in the first, second, and third

**Figure 2.5: Words Most Strongly Predictive of Secrecy in Matched Sample (Procedural Secrecy Analysis)**



*Notes:* For each of the words listed on the lefthand side of the plot, a solid line maps that word to its most likely topic (given estimates from the LDA model described in B.9). A dotted line maps each word to its second most likely topic. The topics listed on the righthand side of the graph are ordered in a specific manner: the uppermost topic is the mode of the modal topic assignments (i.e., the topic that is most frequently the modal topic assignment for the top RF terms), while subsequent topics are presented in descending order according to their similarity to the first topic. Topical similarity determined by the cosine similarity between topic vectors. The plot reveals remarkable concordance on the following: words that are most predictive of secrecy tend to be used in similar topics, and those topics tend to concern the official business of foreign leaders, their meetings, and words relating to information exchange. In general, these words also have positive positive lasso regression coefficient estimates, which implies their use, on the margins, is positively associated with cable restriction.



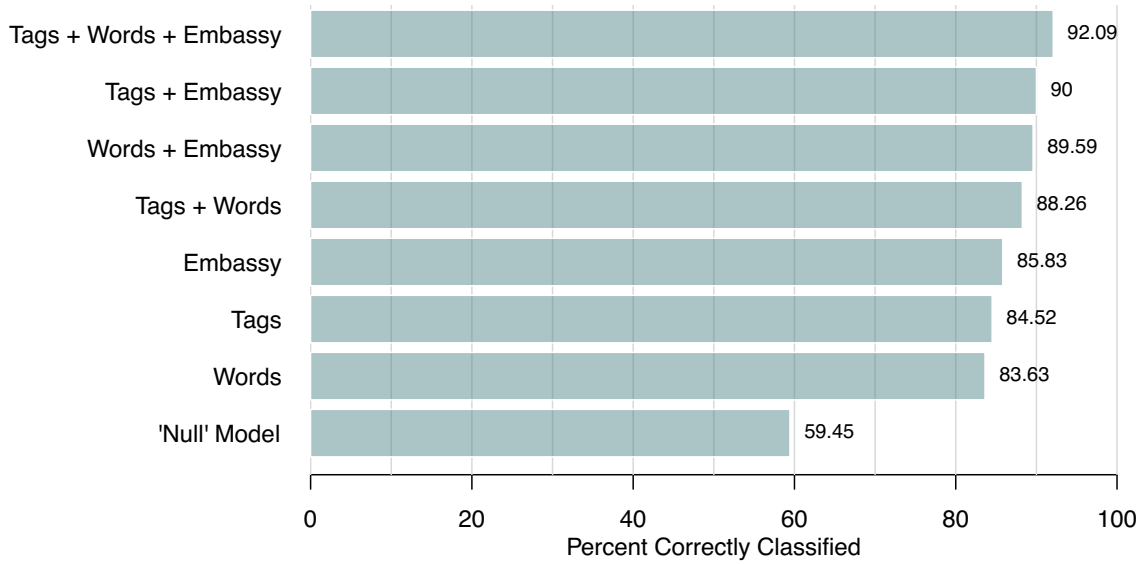
topic as a leading word, while 'will' appears in the fourth topic. Importantly, the words that we have identified as discriminating between unrestricted and restricted cables do *not* appear alongside obviously substantive subject matter such as pertains to the Middle East (topic six or topic seven), the Pacific rim (topic 8), nuclear proliferation (topic 16) or Russian aggression (topic 14). Of course, we do see that some terms are likely to appear within certain substantive topics (such as 'meet', which appears in a 'Burma' topic and 'demarch[e]' which appears in an Israel topic towards the bottom of the plot). Such occurrences are not the norm, however.

In terms of the theories presented earlier, our finding here seems most closely compatible with the work of Sartori (2002) and Kurizaki (2007) insofar as privacy seems to be intrinsically valued by diplomats, rather than because it allows *per se* information exchange.

### 2.5.3 Share of Secrecy: Substance vs. Procedure

Above, we made the claim that while some of observed diplomatic censorship is a consequence of the need to protect state secrets, at least part of it results from the need to keep meetings confidential as a procedural requirement, regardless of what is to be discussed. In our final set of results, we attempt to estimate the relative contribution that these two separate elements make to the practice of restricting information from public view. In Figure 2.6 we report a comparison of models with this in mind. Here, 'Tags' refers to the tag covariates we noted earlier, 'Embassy' are simply embassy fixed effects, and 'Words' are the top 30 words selected by the Random Forest procedure above. In all cases, the numbers to the right of the bars refer to the percent correctly predicted (unrestricted and restricted) by a given (logit) model in the entire sample of 163,958 documents.

**Figure 2.6:** Classification Performance of Models that Incorporate Words, Tags, and Embassy-Level Information



*Notes:* Comparison of classification performance for logistic regression models with 10-fold cross validation on the full post-2005 sample ( $n = 163,958$ ). 'Tags' denote presence of subject tags, 'Words' the top 30 RF words, and 'Embassy' denotes inclusion of embassy IDs. The value of the 'null model' represents the proportion of the modal restriction status in the data set (i.e., restricted).

Unsurprisingly, we see that a model with tags, the word information, and the embassy fixed effects does best in terms of the proportion of documents it can classify correctly, at around 92%. The null model, the sample proportion of restricted cables is 59%, and clearly the statistical model improves substantially upon this. More interesting from our perspective is a comparison of the second and third bar ('Words + Embassy' and 'Tags + Embassy'), and the fifth and sixth ('Words' and 'Tags') since the performance of the models using the RF words and tags are so similar. That is, it seems that whether we use the substantive topics alone, or the words that we identified as connoting secret meetings rather than substance, our model performs similarly. This suggests, at the very least, that both substantive and procedural secrecy matter for diplomatic communication, and that both 'audience cost'-type theories and more recent work on communication have some

empirical support in the data.

## 2.6 Discussion

Conflict and bargaining have always been at the core of international relations and its study (see, e.g., Thucydides, 1910/431BCE; von Clausewitz, 1832/1989). In recent times, the discipline has amassed an impressive array of theoretical models that make use of, or provide findings for, ‘information’ and its dissemination between actors. This paper opened by noting that, despite this voluminous literature, there is little systematic statistical work on the subject, and that this is hardly surprising given that secrets—by definition—are difficult to observe. In this paper, we made use of a unique disclosure of diplomatic communications, a new and contemporary dataset that has an unusually large amount of ‘uncensored’ content (that was not systematically edited), to examine the empirical support for various conceptions of secrecy and communication. We argued that diplomatic confidentiality, i.e. information actively kept from the public, is used in at least two scenarios or ‘dimensions’: first, in a way pertaining to *substance* and second, pertaining to *procedure*. In the former case, documents that deal with issues that could damage U.S. capabilities were they available to others, are disproportionately kept secret. Meanwhile, in cases where publicity is helpful to the U.S. government are made available either for direct public consumption or for distribution to those who will have the opportunity to influence opinion. More speculatively, and untested here, information may be released because it creates a useful ‘audience cost’ and encourages commitment to a costly path of action. In the second case, that of the procedural dimension, diplomats ensure that the circumstances and process of meetings in general—regardless of their actual subject content—are not disclosed. To be clear, we found evidence of both dimensions in our data, and were able to characterize their content and nature. In this way, both the recent literature that emphasizes the importance of diplomacy (e.g., Sartori, 2002; Kurizaki,

2007; Trager, 2010), and extant formal literature that has it playing little role, finds some support here.

Apart from the preliminary analysis our paper provided, it also contributed methodologically to a growing area of political science: that of text analysis. In particular, we faced a situation in which documents had to be compared within particular subject areas, such that their discriminatory terms could be uncovered. We used an exact matching algorithm to estimate textual quantities of interest. In our case, the subject matter was determined by the U.S. State department (via the TAGS system), but the inferential setting is more general. For example, one might be interested in the success (or otherwise) of different bills in Congress or the public opinion reception of speeches from primary candidates. Clearly, the subject matter between documents differs and needs to be 'controlled' for in some sense. We provided one way of proceeding in such situations.

Of course, analytically, we have only scratched the surface here. Though we document the nature and structure of secrecy and the cables themselves, there is much more to do. First, while we argue that 'more secretive' topics in the TAGS system seem to deal more fully with capabilities than the 'least secret', we are necessarily vague on the details. We would like to know more about why exactly some subjects are kept from public view, and whether such decisions accord with IR theory in the area: for example, is topic secrecy actually dictated by a desire to avoid revealing capabilities on a particular subject, or is it more connected to notions of resolve, or even just the information dispensing machinery itself? Second, although we do not directly engage the plausibility of the audience cost literature and its critics (see Slantchev, 2012: for a review), our findings are at least minimally consistent with both sides of that debate, insofar as we find some evidence that the U.S. attempts to make more public its views (and thus possibly create such 'audience costs') where helpful, but not always. That is, it seems to preserve 'room for maneuver' in some areas. Subsequent analysis might weigh in more helpfully on this debate by considering the constraints that U.S. officials face in the various areas

of international relations with which it deals: for example, we might be interested to know whether, in fact, issues that the United States is seemingly 'open' about with the public are simply those where it cannot be otherwise given commonly held knowledge about the U.S. position (or its weaknesses) in the wider world. This is ultimately a call to incorporate more topic-specific covariates and circumstances in the analysis. Finally, while we have emphasized the importance of private diplomatic meetings as part of the arsenal of U.S. international relations practice, we have done little to explain how or why they are used. That is, we are not much the wiser as to which of the various theories (Sartori, 2002; Kurizaki, 2007; Trager, 2010: e.g.) of diplomatic exchange is correct, if any. A continued digging into this corpus might allow for more direct tests of these models. We leave such questions for future work.

# Chapter 3

## A Causal Text Analysis of How Federal Reserve Discussions Respond to Increased Transparency.<sup>1</sup>

### 3.1 Introduction

*“Release of videotape, audiotape, or a literal transcript would have a chilling effect on the free flow of ideas and the ability to bring confidential information to the deliberations.”*

— Alan Greenspan, FOMC Transcript, October 5, 1993

*“Quicker and more complete disclosure already has changed the nature of the Committee’s deliberations. I am for the disclosure that we do, but we should not mislead ourselves about how it has changed the nature of these proceedings. I recall participating in routine, vigorous, and freewheeling debates in this room before we decided to release transcripts. Now, most of us read prepared remarks about our Districts and the national economy and even our comments on near-term policy sometimes are crafted in advance. Prepared statements were the rare exception rather than the rule until we started to release transcripts.”*

— Ed Boehne, President of the Philadelphia Federal Reserve Bank, June 1998 Transcript

We study the effects of transparency on the deliberations of the Federal Reserve Open Market Committee (FOMC). In 1993, the FOMC switched from a regime where their

---

<sup>1</sup>Co-authored with Michael Egedal and Martin Rotemberg. We have benefited greatly from discussions with Miguel Acosta, Jeff Frieden, Claudia Goldin, Rick Hornbeck, Danial Lashkari, Corey Lynch, Ellen Meade, Andrea Prat, Julio Rotemberg, Bryce Millett Steinberg, Paul Tucker, and Fernando Yu, as well as various seminar participants.

meetings were thought to be secret to a regime where it was known that the public could read what was said. If FOMC members care about their public perception, they will they shift their language previously-private deliberations in response to this type of reform, and we show theoretically that they will do so to more closely resemble their public speech. We develop new methods in order to test the theory, and confirm that transparency caused FOMC members to adjust their speech due to a shift to more “public-friendly” language. We also build new computational methods to account for semantic similarity, and show that the observed shift in language was not due to substitution to words with similar meanings.

Text data are well-known to be difficult to use in a systematic and replicable manner, though their merits are clear in a range of scientific settings (e.g., Bloom, Schankerman and Van Reenen, 2013; Gentzkow and Shapiro, 2010; Griffiths and Steyvers, 2004; Roberts et al., 2014). In analyses like ours, for example, there are thousands of words in the English language, which leaves applied researchers with an almost overwhelming number of potential choices to make in the pre-processing of the data (Grimmer and Stewart, 2013a). Furthermore, different words often have related meanings: just because two documents use literally different words does not mean that they are truly about different topics or issues.<sup>2</sup> To overcome these obstacles, a variety of strategies have been developed in the quantitative text analysis literature, which range from analysis on words identified ex-ante to using modern clustering methods and “latent variable” models (e.g., Blei, Ng and Jordan, 2003) to reduce the dimensionality of the research setting.

One disadvantage of relying solely on these types of methods is that they makes replication across contexts difficult. For example, this paper uses language from the central bank of the United States, and we would like to generate results which in principle could be compared to the effects of the recent switch to more transparent deliberations at

---

<sup>2</sup>Along the same lines, words that appear to be the same may in fact have different meanings in different contexts (e.g., Navigli, 2009; Turney and Pantel, 2010).

the Bank of England. If we were to use within-FOMC documents to generate measures or word relations, which would follow the standard practice in the literature, then it would be challenging to compare those results to an analysis which instead generated clusters using within-Bank of England language.<sup>3</sup> In our application, we address this issue through use of an external source—a technical dictionary—to generate an out-of-sample measure of the relationship between words in our data.<sup>4</sup> To analyze aggregate and word-level results, the method we introduce has the added benefit of not requiring ex ante dimension reduction to “topics” or clusters, since in many contexts it is rarely clear what the appropriate number of latent categories is (e.g., Grimmer and Stewart, 2013a; Wallach, Mimno and McCallum, 2009; Wallach et al., 2009).

The main set of documents we study come from FOMC deliberations and public documents. Starting in 1976, archivists at the Federal Reserve kept recordings of FOMC meetings, without the knowledge of most of the participants. The Federal Reserve publicly denied the existence of those recordings (Auerbach, 2011)—most FOMC members did not know of their existence—and the only public information about the meetings were short summaries. When the existence of the recordings was revealed in 1993, the Federal Reserve agreed to release all of the transcripts from the earlier meetings. Furthermore, the FOMC agreed to continue to release summaries soon after meetings, and to begin releasing full meeting transcripts with a five year lag—an arrangement which still continues.

Each meeting corresponds to one public summary and one transcript. Comparing the documents over time allows us to uncover the effect of transparency, since the summaries have always been public, and the deliberations were only known to be transparent after the policy change. For our comparison, we start with a commonly-used vector

---

<sup>3</sup>Strikingly, even if we used a slightly different sample of FOMC documents, we would generate different clusters of words.

<sup>4</sup>Of course, our estimation approach is not incompatible with estimates of word relations derived from the raw FOMC meetings themselves; but for purposes of replicability, due to the particularity of words in economic speech, and to address concerns of a possible language shift to words with similar meanings, we prefer to scale word relations through an outside source.



space measure known as “cosine similarity” (e.g., Salton and McGill, 1986; Hoberg and Phillips, 2010, 2016; Kang, 2015), which measures the cosine of the angle between two  $n$ -dimensional vectors. In our context, each word is a separate dimension.<sup>5</sup> We show that this approach has micro-foundations in a straightforward model of career concerns, and that from an applied perspective the cosine similarity measure has several desirable properties. We find that the similarity of the private and public documents increased by around 20% after the policy change.

Part of the impetus for the development of modern clustering methods is the difficulty in interpreting that 20% number, since it cannot be used to discover whose behavior changed, nor how. For instance, the change could come from the FOMC members behaving differently in their deliberations, or from different behavior when summarizing the deliberations in the public documents. Even if the deliberations are the documents which change, the reasons for the change could be either a decrease in irrelevant discussions (such as a shift away from small talk), or a shift in substantive language. Furthermore, since different words have overlapping meanings, if the change in similarity were driven by a shift in language to describe the same issue—such as switching from “currency” to “money”—then the explanation of the effect of transparency would be more about organizational behavior than about a substantive change in focus. These concerns are present for most settings where similarity is the object of interest. For instance, when studying dynamics of racial segregation over time, it would be valuable to know both which groups’ behavioral change led to changes in segregation, and to account for potential overlap between groups.

We would like to develop a theory to explain the response to the policy change, and in order to do so we develop methods to address both of these concerns, that should be useful to other researchers interested in similarity. We develop a first-order decomposition

---

<sup>5</sup>Two documents would have a cosine similarity of 0 if they contained no overlapping words, and a similarity of 1 if every word’s proportions were the same in each document.

from the change in similarity over time into changes in behavior at the word-document level. We show that each word's contribution to the growth in similarity is the product of two terms: the word's own growth over time, and a term which measures the "gap" between that word's usage in each document in the pre-period. For a given gap, increasing a word's growth over time will increase the magnitude of that word's contribution to the change in aggregate similarity. The change in similarity is fundamentally related to the covariance between the growth in the usage of each word and its associated gap. We also describe a family of similarity measures which includes both the cosine similarity measure and the commonly used Atkinson index, which we use to show that the change in similarity is not driven by sparse words.

The decomposition we develop has the useful feature that it does not require specifying particular words in advance: it uncovers each word's contribution to overall changes in document similarity, and therefore can create rankings of the words which are most responsible for these differences. This allows for dimension-reduction—a common and important feature of most quantitative text analysis methods, which study data with a relatively large number of covariates. However, unlike the standard in the literature, we do not need a pre-processing step to lower the dimensionality, either by creating clusters in the data or using ad-hoc methods to identify ex-ante "interesting" words or phrases. We find that the increase in similarity after the policy change is driven by specific language choices in the deliberations. Furthermore, very few of the words—around 5 percent—are responsible for 90 percent of the change in similarity after transparency. The words most responsible for the change tend to be economically meaningful, such as "inflation" and "growth." Furthermore, words that are commonly used to convey personal opinions and thoughts, such as "think" and "say"—were used substantially less in the meetings following transparency reforms, consistent with qualitative evidence that FOMC members began to prepare speeches after the transparency reform.

To account for words' overlapping meanings, we leverage the fact that the purpose

of technical dictionaries is to relate words to each other through definitions. We use the definitions of words in the Oxford Dictionary of Economics to create a measure of how similar words' meanings are to one another.<sup>6</sup> The ODE allows us to develop the meaningful measures of semantic similarity for the kind of technical language utilized in Federal Reserve Discussions.<sup>7</sup> We show how to adapt standard measures of similarity, which traditionally treat each dimension (or word) as orthogonal, to account for relational weights. This allows us to distinguish a change in word choice from an actual change in content. The increase in similarity of the private and public texts after transparency reforms remains even when accounting for words' meanings.

The mechanisms why transparency would lead to a substantive change in language use at the previously-private deliberations can be captured by adapting relatively standard models of career concerns. As a result, the welfare consequences of the behavioral change are ambiguous, since they may be driven by either a reallocation of effort or an increase in overall effort. Our result is consistent with the mixed predictions in the existing theoretical literature on transparency.<sup>8</sup>

There are two contemporaneous projects who study the same transparency reform. Acosta (2015) also uses cosine similarity to study the effects of the transparency (and finds similar results to our Table 1), but focuses on within-meeting interactions. Woolley and Gardner (2009) develop a language-based measure of deliberation, and find that it declines after the policy reform. Hansen, McMahon and Prat (2014) revisit Meade and Stasavage (2008) by first generating "topics" in the FOMC transcripts, and discussing topic usage over time, and focus on how FOMC members' speech relates to each other,

---

<sup>6</sup>See Lesk (1986) for another approach for using dictionaries to estimate for semantic relations. Other strategies for generating measures of semantic similarity include using Wordnet (Miller et al., 1990; Resnik, 1995; Miller, 1995) or a thesaurus (McHale, 1998).

<sup>7</sup>This strategy is applicable to a wide variety of computational linguistics settings beyond economic policy discussions, as there are over 300 Oxford Reference Dictionaries, including ones for medical, legal, and musical language.

<sup>8</sup>See, for example, Fama (1980), Holmstrom (1979), Dranove et al. (2003), and Prat (2005).

essentially a sophisticated study of nuanced peer effects. However, as is well-known in the peer effects literature (Manski, 1993; Angrist, 2014), changes in the relationship between agents' outcomes may be due to a common external shock. Our results show that this is an important consideration for understanding this policy reform, since we observe a secular shift towards more "public-friendly" language in the deliberations after the policy change.

Our paper contributes to a growing literature studying FOMC deliberations. Swanson (2006) argues that the shift to FOMC transparency in the 1990s, which includes the release of the transcripts but also includes a variety of other changes, such as in February 1994 deciding to start explicitly announcing changes in the federal funds rate target, led to markets being less surprised by FOMC actions. We complement this result by focusing our empirical efforts on the effects of the policy change on deliberations. Many authors have used FOMC transcripts in order to understand other trends in the economy.<sup>9</sup> Furthermore, a growing body of research discusses effects of central bank transparency reforms.<sup>10</sup>

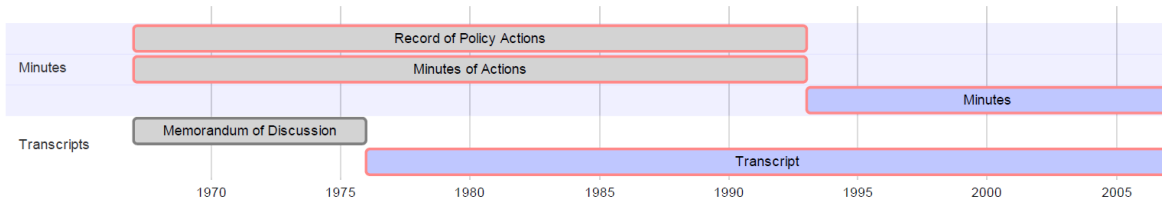
The structure of this paper proceeds as follows. In Section 3.2, we describe the historical context of the policy and our data sources. In Section 3.3, we describe the core similarity metric used in our analysis, its word-level decomposition, and our empirical strategy. In Section 3.4, we show our empirical estimates of the effects of transparency, and Section 3.5 concludes.

---

<sup>9</sup>See, for example, Romer and Romer (2004), Meade and Thornton (2012), Schonhardt-Bailey (2013), Rotemberg (2013), and Fligstein, Brundage and Schultz (2014).

<sup>10</sup>See, for example, Geraats (2002), and Bernanke and Kuttner (2005)

**Figure 3.1: Changing Naming Conventions of Federal Reserve documents from 1967–2007**



*Notes:* Our study uses documents from 1976 to 2007. We sometimes generically refer to the summaries as “minutes” and the detailed internal information as “transcripts.” This figure shows the underlying names that were contemporaneously used by the FOMC.

## 3.2 Historical Background

### 3.2.1 Data availability

The Federal Open Market Committee, formed in 1935, has publicly released “Records of Policy Actions” for most of its existence; these were at first released only once a year. The Committee also maintained private records called “Minutes,” which contained, for each meeting, details on attendance, discussions, and decisions. In 1967, the Records of Policy Actions started to be released roughly ninety days after each meeting. The Minutes were split into two parts, with the new second document, called the “Minutes of Actions,” made available to the public; the other document, called the “Memorandum of Discussion,” was kept private. The delay on the release of the public documents was further cut to 45 days in 1975. Our data starts in 1976, when the delay was decreased further to 30 days (although this deadline was not always met).

These summaries were called “Records of Policy Action” prior to 1993, and “Minutes” thereafter.<sup>11</sup> The official names of the documents over time are presented in Figure 3.1.

<sup>11</sup>Despite having different names, Records of Policy Action and Minutes are “functional equivalents,” according to the Federal Reserve. See, for example, Danker and Luecke (2005). The results of our later analysis, where we decompose observed changes in aggregate similarity between documents, provide further evidence to this claim.

In 1976, Congress passed the Government in the Sunshine Act which said that government agencies “shall make promptly available to the public, in a place easily accessible to the public, the transcript, electronic recording or minutes of the meeting.”<sup>12</sup> In a 10–1 vote, the FOMC voted to discontinue the keeping of transcripts, to make it impossible to release any information publicly (Auerbach, 2009).

In 1993, the House Banking Committee, led by former United States Representative Henry B. Gonzalez (D-Texas), discovered that recordings of the meetings existed. An agreement was reached where the FOMC would release lightly edited transcripts with a five-year lag. The new transparency rules were recognized in the popular press (Friedman and Schwartz, 1993), and it became well-known to members of the FOMC that the transcripts would be read by critical citizens.

There are generally 19 FOMC members, who meet 8 times a year (4 meetings is the statutory minimum). There is a chairperson, who typically serves for about a decade, and 6 other Governors based in Washington, DC. Furthermore, there are 12 regional banks, who send their President to the meetings. Although all of the members speak, only a subset of the regional presidents have voting power.<sup>13</sup> While being on the FOMC is a highlight of any career, some of the members are only there for a short period of time (many academics only serve until their universities threaten to pull their tenure), while others end up advancing through the Fed—all of the FOMC chairpersons during the authors’ lifetimes were first on the FOMC. While there were several FOMC members whose tenure spanned both sides of the transparency reform, we do not constrain our analysis to just those members.

---

<sup>12</sup>We have been informed by Federal Reserve employees that the institution’s belief is that they are exempt from this legislation, although this question has not been settled by the courts.

<sup>13</sup>The President of the New York Fed always has a vote, the Chicago and Cleveland presidents each vote every other year, and the rest of the members rotate in to vote for one out of every three years.

### 3.2.2 Data

The data for our analysis consist of 268 publicly available transcripts from 1976 to 2007 with their corresponding public summaries.<sup>14,15</sup>

## 3.3 Methodology

We follow the standard in the text analysis literature by using a “bag of words” approach. We convert all transcripts and public summaries into vectors of word counts; then we “stem” the words and remove “stop” words.<sup>16</sup> Having generated the vectors, we introduce measures for describing the effect of a policy change on language choice. In particular, we leverage the two types of documents released by the FOMC, and focus on how similar those documents are to each other, and how that relationship evolves over time. We begin by describing properties that we would like a similarity metric to satisfy, and then propose a particular metric, “generalized cosine similarity,” which is the first method (that we know about) that satisfies those properties.

### 3.3.1 Notation

Following standard terminology, we let  $w \in \mathcal{D} = \{w^1, \dots, w^D\}$  denote a *word* for a given dictionary  $\mathcal{D}$ . Over the  $T$  periods in the data, there are  $2T$  *documents* in the *corpus*, with  $\mathcal{C} = \{p_1, \dots, p_T, q_1, \dots, q_T\}$ . We let  $n_{pt}^j$  as the number of times  $w^j$  appears in  $p_t$ , with the corresponding meaning for  $n_{qt}^j$ . Therefore,  $p_t = (n_{pt}^1, \dots, n_{pt}^D)'$  denotes the *document-term vector* for document  $p_t$ .

Define  $\Omega$  to be a  $D \times D$  matrix capturing the relationship of words for a given

---

<sup>14</sup>See <http://www.federalreserve.gov/monetarypolicy/fomchistorical.htm>

<sup>15</sup>We used the OCR software ABBYY FineReader to convert these documents into machine readable text files.

<sup>16</sup>We use the set of Snowball stop words, and the Porter (1980) stemming algorithm.

dictionary  $\mathcal{D}$ , where  $\omega^{ij} \in [0, 1]$  is the *relationship* between words  $w^i$  and  $w^j$ . As words become more related,  $\omega^{ij}$  is increasing; in particular,  $\omega^{ii} = 1$ . We impose symmetry on the relationship between words, so that  $\omega^{ij} = \omega^{ji}$ .  $R_t^{\mathcal{D}\Omega}$  is the *similarity metric* between documents  $p_t$  and  $q_t$  using dictionary  $\mathcal{D}$  and relationship matrix  $\Omega$ . We drop the superscripts for clarity except when they are needed.

### 3.3.2 Similarity Metric Axioms

In this section we present desirable axioms for a similarity metric to satisfy, in the spirit of Tversky (1977), Frankel and Volij (2011), and Bloom, Schankerman and Van Reenen (2013). In addition to the properties of a distance metric,<sup>17</sup> we would like:

1. Addition: Consider two periods, where  $R_t^{\mathcal{D}\Omega} \geq R_{\hat{t}}^{\mathcal{D}\Omega}$ . If we add a word  $w^{D+1}$ , not contained in any of the relevant documents, to our dictionary to form  $\mathcal{D}'$ , then  $R_t^{\mathcal{D}'\Omega} \geq R_{\hat{t}}^{\mathcal{D}'\Omega}$ . This is to say, document similarity should not be a function of the dimensions that they commonly do not have, so constraining our dimension space to the words which appear in the documents (instead of all words that have ever appeared in any language) should not change similarity rank.
2. Monotonicity:  $\forall \alpha_i, \alpha_j > 0$ , if  $p_{t'} = \alpha_i p_t + \alpha_j q_t$  and  $q_t = q_{t'}$ , then  $R_{t'} \geq R_t$ , with equality only if  $R_t = 1$ . Moving “towards” the comparison document should increase the underlying similarity.
3. Synonym invariance: Suppose  $q_t = q_{t'}$  and  $p_{t'}$  and  $p_t$  are identical but for words  $w^k$  and  $w^\ell$ . If  $n_{p_t}^k + n_{p_t}^\ell = n_{p_{t'}}^k + n_{p_{t'}}^\ell$  and  $\omega^{kd} = \omega^{\ell j} \forall j \in [1, D]$ , then  $R_{t'} = R_t$ . When comparing two documents, replacing one word with its synonym should not change the similarity.

---

<sup>17</sup>Non-negativity, so  $R_t \in [0, 1]$ ; identity, so if  $p_t = q_t$ , then  $R_t = 1$ ; symmetry; and the triangle inequality.



4. Within-word similarity: If  $q_t = q_{t'}$ , and  $p_{t'}$  and  $p_t$  are identical but for word  $w^k$ , then  $R_t < R_{t'}$  iff  $\|p_t - q_t\| > \|p_{t'} - q_{t'}\|$ . If the distance for all word but one is held constant, then increasing distance along that dimension should correspondingly decrease similarity.
5. Cross-word similarity: Suppose we have two weight matrices  $\Omega$  and  $\Omega'$  which are identical but for  $\omega^{kl} > \omega'^{kl}$ , and  $R_t^{D\Omega'} < 1$ . This implies that  $R_t^{D\Omega} \geq R_t^{D\Omega'}$ , with equality iff  $\min \{n_{p_t}^k, n_{p_t}^\ell\} + \min \{n_{q_t}^k, n_{q_t}^\ell\} = 0$ . This is the analogue for within-word similarity, but taking into account semantic similarity.

### 3.3.3 Cosine Similarity and Generalized Cosine Similarity

#### Cosine Similarity

For two document-term vectors  $p_t$  and  $q_t$ , the *cosine similarity* of  $p_t$  and  $q_t$  is defined as

$$CS(t) = \frac{\langle p_t, q_t \rangle}{\|p_t\| \cdot \|q_t\|},$$

where  $\langle \cdot \rangle$  represents the dot product and  $\|\cdot\|$  represents the Euclidean norm.

This similarity metric is simple to calculate and satisfies all axioms other than cross-word similarity and synonym invariance. We present a brief proof for Monotonicity, as the rest are straightforward. If  $p_{t'} = \alpha_i p_t + \alpha_j q_t$  and  $q_t = q_{t'}$ , then  $CS_{t'} = \frac{\langle \alpha_i p_t + \alpha_j q_t, q_t \rangle}{\|\alpha_i p_t + \alpha_j q_t\| \cdot \|q_t\|}$ . By the triangle inequality, this is weakly larger than  $\frac{\alpha_i \langle p_t, q_t \rangle + \alpha_j \langle q_t, q_t \rangle}{\|\alpha_i p_t + \alpha_j q_t\| \cdot \|q_t\|}$ , which in turn is weakly larger than  $CS(t)$ , and strictly larger if  $CS(t) < 1$ .

#### Generalized Cosine Similarity

The *generalized cosine similarity* in period  $t$  is defined as

$$CS_\Omega(t) = \frac{\langle p_t, q_t \rangle_\Omega}{\|p_t\|_\Omega \cdot \|q_t\|_\Omega},$$

where now the dot product and the norm are in  $\Omega$  space.<sup>18</sup>

For symmetric, nonnegative, positive definite  $\Omega$ , this similarity metric satisfies all of the axioms. We present a sketch of the proof for synonym invariance, as the rest are straightforward. Suppose  $q_t = q_{t'}$  and  $p_{t'}$  and  $p_t$  are identical but for words  $w^k$  and  $w^\ell$ , and  $n_{pt}^k + n_{pt}^\ell = n_{p_{t'}}^k + n_{p_{t'}}^\ell$  and  $\omega^{kd} = \omega^{\ell d} \forall d \in [1, D]$ . To show that  $CS_\Omega(t) = CS_\Omega(t')$ , we first show that  $\langle p_t, q_t \rangle_\Omega = \langle p_{t'}, q_t \rangle_\Omega$ . Since  $\langle p_t, q_t \rangle_\Omega = \langle (p_t \Omega)', q_t \rangle$ , it is sufficient to show that  $p_t \Omega = p_{t'} \Omega$ . This is easily verified algebraically.

This implies that  $\langle p_t, p_t \rangle_\Omega = \langle p_{t'}, p_t \rangle_\Omega = \langle p_{t'}, p_{t'} \rangle_\Omega$ . As a result, both the numerators and denominators of  $CS_\Omega(t)$  and  $CS_\Omega(t')$  are identical, which completes the proof.

### Extensions of Cosine Similarity

There are many measures of similarity. A commonly used measure, the symmetric Atkinson Index (Atkinson (1970), Frankel and Volij (2011)), is

$$A(t) = \sum_{j=1}^D \left( \frac{n_{pt}^j}{\sum_{k=1}^D n_{pt}^k} \right)^{\frac{1}{2}} \left( \frac{n_{qt}^j}{\sum_{k=1}^D n_{qt}^k} \right)^{\frac{1}{2}}.$$

We can rewrite this measure<sup>19</sup> as the cosine similarity of  $p_t^{\frac{1}{2}}, q_t^{\frac{1}{2}}$ , where  $p_t^{\frac{1}{2}}$  denotes taking the element-by-element square root of the  $p$  vector.<sup>20</sup> We can further extend the cosine similarity as

$$ECS(t, m) = (p_t^m, q_t^m)$$

where  $m$  mediates the weight on more-populated elements. The extended cosine similarity measure is useful because it allows for a diagnostic on the role that sparse words play in

<sup>18</sup>So, for instance,  $\langle p_t, q_t \rangle_\Omega = p_t \cdot \Omega \cdot q_t$ .

<sup>19</sup>The Atkinson index is often used to study inequality, and therefore written as  $1 - A(t)$  so that an increase in the index indicates more inequality instead of more similarity.

<sup>20</sup> $A(t) = \frac{\sum_{j=1}^D p_j^{\frac{1}{2}} q_j^{\frac{1}{2}}}{\left(\sum_{j=1}^D p_j\right)^{\frac{1}{2}} \left(\sum_{j=1}^D q_j\right)^{\frac{1}{2}}}$ .

the evolution of similarity. If the growth of similarity is driven by relatively commonly used words, then the extended cosine similarity will be increasing in  $m$ . At the extreme, if there is no movement on the extensive margin then there will be no change in the extended cosine similarity for  $m = 0$ .

### 3.3.4 Growth in Cosine Similarity

In much of our analysis, we focus not only on the overall similarity of the public and private documents, but also decompose the change in similarity into word-level changes. In particular,  $\widehat{CS}(t) = \langle \widehat{p}_t, \widehat{q}_t \rangle - \|\widehat{p}_t\| \cdot \|\widehat{q}_t\|$ . Some algebra yields

$$\begin{aligned} \widehat{CS}(t) &= \sum_{j=1}^D \widehat{n}_{pt}^j \left[ \frac{n_{pt}^j \cdot n_{qt}^j}{\langle p_t, q_t \rangle} - \frac{(n_{pt}^j)^2}{\|p_t\|^2} \right] \\ &\quad + \sum_{j=1}^D \widehat{n}_{qt}^j \left[ \frac{n_{pt}^j \cdot n_{qt}^j}{\langle p_t, q_t \rangle} - \frac{(n_{qt}^j)^2}{\|q_t\|^2} \right]. \end{aligned} \quad (3.1)$$

For generalized cosine similarity,

$$\begin{aligned} \widehat{CS}(t)_\Omega &= \sum_{j=1}^D \widehat{n}_{pt}^j \left[ \frac{n_{pt}^j \cdot \sum_{i=1}^D \omega_{ij} n_2^i}{\langle p_t, q_t \rangle_\Omega} - \frac{(n_{pt}^j) \cdot \sum_{i=1}^D \omega_{ij} n_1^i}{\|p_t\|_\Omega^2} \right] \\ &\quad + \sum_{j=1}^D \widehat{n}_{qt}^j \left[ \frac{n_{qt}^j \cdot \sum_{i=1}^D \omega_{ij} n_1^i}{\langle p_t, q_t \rangle_\Omega} - \frac{(n_{qt}^j) \cdot \sum_{i=1}^D \omega_{ij} n_2^i}{\|q_t\|_\Omega^2} \right]. \end{aligned}$$

Similarly, for extended cosine similarity,

$$\widehat{CS}(t, m) = \sum_{j=1}^D m \cdot \widehat{n}_{p_t}^j \left[ \frac{(n_{p_t}^j \cdot n_{q_t}^j)^m}{\langle p_t^m, q_t^m \rangle} - \frac{(n_{p_t}^j)^{2m}}{\|p_t^m\|^2} \right] + \sum_{j=1}^D m \cdot \widehat{n}_{q_t}^j \left[ \frac{(n_{q_t}^j \cdot n_{p_t}^j)^m}{\langle q_t^m, p_t^m \rangle} - \frac{(n_{q_t}^j)^{2m}}{\|q_t^m\|^2} \right].$$

As a result, the change in similarity can be decomposed into the growth rates of the respective words.<sup>21</sup> Define  $w_{p_t}^j \equiv \frac{n_{p_t}^j \cdot n_{q_t}^j}{\langle p_t, q_t \rangle} - \frac{(n_{p_t}^j)^2}{\|p_t\|^2}$ , with an equivalent definition for  $w_{q_t}^j$ . This corresponds to the “gap” in usage of word  $j$  between documents  $p_t$  and  $q_t$ .  $w_{p_t}^j$  is positive if and only if word  $j$  is relatively overrepresented in document  $q_t$ . As a result, if and only if  $w_{p_t}^j$  is positive will increasing the usage of word  $j$  in document  $p_t$  increase  $CS(t)$ . In other words, if a word’s use in a document increases, this leads to an increase in similarity if and only if the word had previously been relatively underrepresented in that document. The magnitude of the effect is increasing in the magnitude of the under-representation.

For any similarity mapping, it is theoretically possible to run numerical counterfactual simulations to discuss the effect that each word’s evolution has on the aggregate change. The cosine similarity measures have the desirable property that, the word-level derivatives are analytically straightforward, relate to a clear intuition, and would be an exact decomposition in continuous time.

### Covariance of Gap and Growth

The total growth of similarity is fundamentally related to the covariance of the growth rate of each of the words and their respective gaps. In particular, some algebra yields

$$\widehat{CS}(t) = D \cdot (\text{Cov}(\widehat{n}_{p_t}, w_{p_t}) + \text{Cov}(\widehat{n}_{q_t}, w_{q_t})). \quad (3.2)$$

<sup>21</sup>For notational clarity, we focus on the Cosine Similarity measure; the intuition is the same for the other measures.

where  $D$  is the total number of words in the dictionary. Multiplying the growth rates of the words therefore will have no effect on the growth rate of similarity (given that cosine similarity is unitless, this is intuitive). The only way to increase the growth rate is for the growth rate of the individual words to positively covary with their ex-ante gaps.

### Growth in Cosine Similarity in the Data

In the data, we are less interested in growth from period to periods, and more interested in the growth from the pre to post transparency regime. We therefore approximate the growth rates in similarity (which in Equation 3.1 is in continuous time) using a discrete approximation following Davis and Haltiwanger (1992). In particular, for private statements  $p$  and press releases  $q$ , we calculate

$$\begin{aligned} \widehat{CS}(t) &\approx \frac{2 [CS(t) - CS(0)]}{CS(t) + CS(0)} \\ &\approx \sum_{j=1}^D \left( \frac{2 (n_{pt}^j - n_{p0}^j)}{n_{pt}^j + n_{p0}^j} \right) [w_{p0}^j] \\ &\quad + \sum_{j=1}^D \left( \frac{2 (n_{qt}^j - n_{q0}^j)}{n_{qt}^j + n_{q0}^j} \right) [w_{q0}^j]. \end{aligned} \quad (3.3)$$

This allows us to determine the effects of transparency for each word-document dyad;  $q_0$  and  $p_0$  are calculated as the average word shares for the respective documents in the pre-period, and we explore robustness to different windows.

Gentzkow, Shapiro and Taddy (2015) argue that analytic methods for measuring similarity are biased downwards when there are many covariates and relatively short documents. In Appendix C.4, we find that those concerns are present in our context, but do not drive our results.<sup>22</sup> In Figures C.28 and C.29, we present evidence that our results are not being driven by the finite lengths of the documents.

---

<sup>22</sup>One important difference may be that they study over 700,000 bigrams, whereas we focus on 4,032 unigrams.

### 3.3.5 Constructing the Term-Relationship Weight Matrix

We also want to see if changes in language use are due to fundamental changes in meaning. We do this by developing a new method for measuring semantic similarity—a specialized *dictionary-based* approach—as a natural successor to conceptual word lists commonly used in psychology and linguistics applications (e.g., Stone and Hunt, 1963). Instead of relying on experts to categorize words as being part of certain topics or relating certain sentiments, we use words’ dictionary definitions in order to determine how they relate to each other. We use the Oxford Dictionary of Economics, which contains 3,423 entries (Since many of the entries are *n*-grams, there are 4,798 corresponding stems. 4,032 of them appear in the FOMC documents).<sup>23</sup> An alternative approach could be to generate topics in the dictionary using the Latent Dirichlet Allocation (Blei, Ng and Jordan, 2003), which we discuss further in Appendix C.2. For example, the entry for *inflation* is

A persistent tendency for nominal prices to increase. Inflation is measured by the proportional changes over time in some appropriate price index, commonly a consumer price index or a GDP deflator. Cost inflation is started by an increase in some element of costs, for example the oil price explosion of 1973–4. Demand inflation is due to too much aggregate demand. Once started, inflation tends to persist through an inflationary spiral, in which various prices and wage rates rise because others have risen. Hyperinflation is extremely rapid inflation, in which prices increase so fast that money largely loses its convenience as a medium of exchange.

The intuition behind our approach is that words which appear in the same entries are likely to be similar. For instance, *price index* is used to define *inflation*, and vice-versa. Furthermore, both are used to define *income*. In order to construct our weight matrix, we start with the ODE document-term matrix, where each entry is a row and the columns consist of all of the words used in any definition<sup>24</sup>.  $\Omega_{ij}$  is constructed by taking the cosine

---

<sup>23</sup>To make our Cosine and Generalized Cosine similarity results directly comparable, we normally use only those 4,032 stems in all of our analysis. Including all of the FOMC stems for the Cosine similarity barely changes the results.

<sup>24</sup>A word is considered part of its own definition.

similarity of columns  $i$  and  $j$  (so it is symmetric by construction).<sup>25</sup>

## 3.4 Results

In this section, we discuss the effects of the transparency reform on the similarity of the transcripts and public documents. For the most part, we do so graphically, plotting over time how similarity evolves.

### 3.4.1 The Evolution of Language after Transparency

A natural starting place for observing the effect of transparency is to compare the transcripts with the public documents over time, using the similarity measures derived in the previous sections. The two types of documents becoming more similar over time is evidence for an effect of the transparency reform. The left panel of Figure 3.2 shows that the similarity of the deliberations and their corresponding public summaries increased following the unexpected enforcement of the Sunshine Act. The middle panel uses only the stems which match to an entry in the dictionary.<sup>26</sup> In the right panel, we incorporate semantic similarity using our term-relationship weight matrix, and, while there is a slightly higher overall similarity, we again see a rise in similarity following late 1993. This suggests that the effect of the policy reform was not merely through language choice.

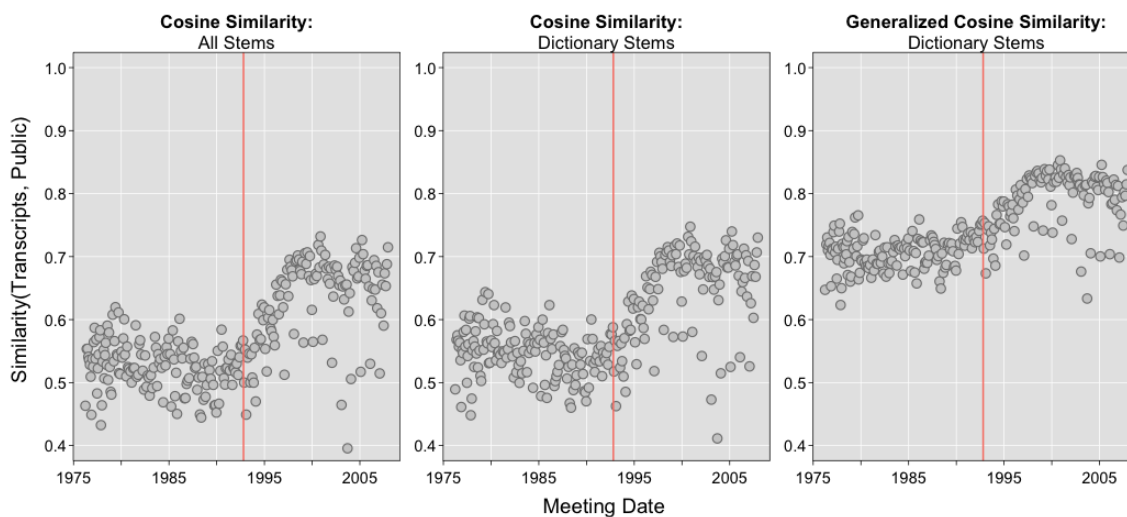
In all cases, the increase in similarity after the transparency reform is not immediate, but gradual, taking several years before reaching a new steady state. The transcription process changed after the reform, but the gradual change contradicts that the observed effect is due to changes in measurement. The gradual increase is, for instance, consistent

---

<sup>25</sup> $\Omega$  is not constructed to be positive-definite, but that is needed to bound the Generalized Cosine Similarity. For instance, one concern is Sylvester's criterion for positive-definiteness, which requires that all of the off-diagonal entries must be less than one. To solve this issue, we rescale the off-diagonal elements by dividing by ten.

<sup>26</sup>The correlation in meeting similarity for all of the words or just the ones which match the dictionary is 0.998.

**Figure 3.2:** Cosine similarity of FOMC transcripts and corresponding public summaries from 1976–2007



*Notes:* This figure plots the similarity of each period’s FOMC Transcript and its associated Public Summary, where each dot corresponds to one period. The first panel includes all word stems in the raw data (a total of 34,616 stems), the second panel includes the subset of stems that appear at least once in the Dictionary of Economics (4,032 stems), and the third panel uses the Generalized Cosine Similarity measure, using the dictionary definitions as weights (4,032 stems). The vertical line corresponds to the timing of the policy change.



with a transition path, as the FOMC may not have immediately known its preferred response to transparency.

The increase in similarity per se is not informative about its underlying cause. One explanation could be that the meeting changed, as FOMC members adjusted their language to be more public-friendly, in the same way that the press releases were always designed with the public in mind (while the press releases remained unchanged). An unrelated story would be that the discussions stayed the same, but that the press release had previously not been a complete representation of the FOMC discussions and changed after it became ex-post verifiable. A well-known issue with vector similarity has been a researcher's inability to distinguish between these two types of stories, which we overcome with the decomposition derived in the previous section.

### Estimating Treatment Effects

We explore several models in order to estimate the “effect” of the policy change on document similarity. It is clear from Figures 3.2 and 3.5 that the FOMC transitioned to a new equilibrium after the policy reform, but for the most part we abstract from this, and focus on the difference before and after the transparency reform.

For our main results, we focus on the first-order approximation to similarity growth, following Equation 3.3. In Table 3.1 we explore how the size of the effect varies as we adjust the sample window (which documents before and after are used to estimate the effect of the policy) and the baseline (which documents before the policy change are used to measure the growth rates of each word). Each cell represents a different regression of the form

$$\hat{S}_{\text{approx},t} = \beta_0 + \beta_1 \text{Post}_t + \epsilon_t$$

reporting both the effect of post reform, as well as the standard error of the estimated effect. Figure 3.5 corresponds to both the sample and the benchmark window of  $\{1 - 50\}$ ,

**Table 3.1: Effect of Transparency on Document Similarity**

Benchmark Window	Cosine Similarity			Generalized Cosine Similarity		
	Sample: {1 – 25}	Sample: {1 – 50}	Sample: {25 – 75}	Sample: {1 – 25}	Sample: {1 – 50}	Sample: {25 – 75}
{1 – 25}	0.115 (0.031)	0.221 (0.038)	0.333 (0.015)	0.053 (0.015)	0.113 (0.018)	0.182 (0.008)
{1 – 50}	0.122 (0.034)	0.235 (0.041)	0.35 (0.015)	0.057 (0.016)	0.122 (0.022)	0.198 (0.009)
{25 – 75}	0.109 (0.037)	0.214 (0.04)	0.311 (0.015)	0.056 (0.018)	0.12 (0.022)	0.196 (0.009)

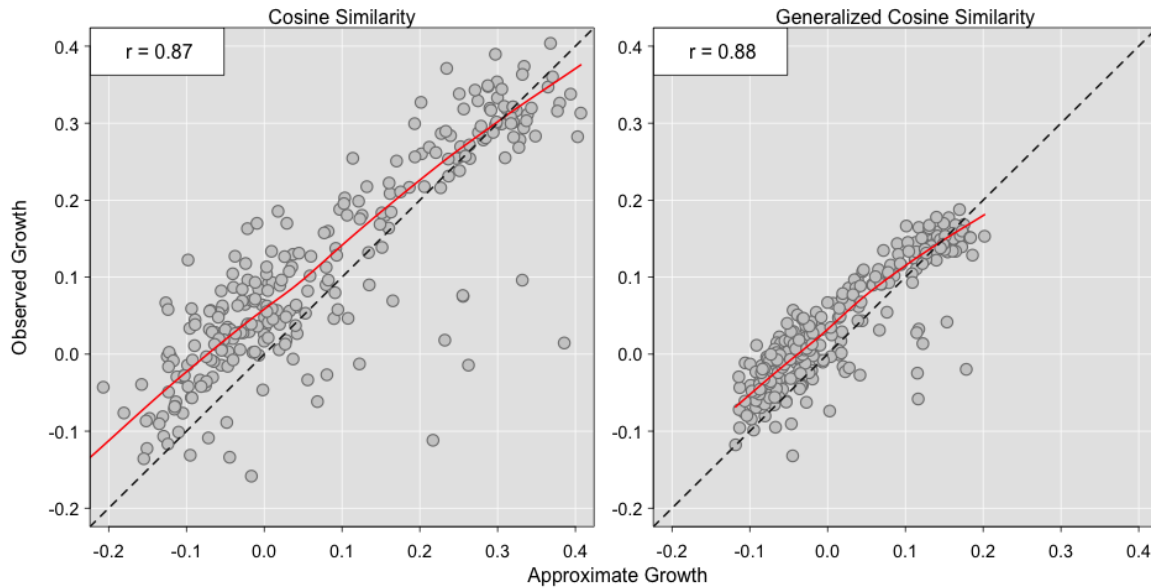
*Notes:* OLS estimates of the effect of transparency on the similarity of FOMC deliberations and public summaries. The various columns represent different samples, they are always symmetric (so {25 – 75} means the meetings 25<sup>th</sup> to 75<sup>th</sup> meetings both before and after the policy change). Each row represents a different benchmark for determining the growth rate. The outcome is the approximated growth rate from Equation 3.3. Newey and West (1994) standard errors in parenthesis.

which is the specification we use for the rest of the picture. As is consistent with the transition path shown in Figure 3.5, increasing the sample for longer after the policy change also increases the estimated treatment effect. However, the estimated coefficients are insensitive to the choice of benchmark window. All of the coefficients are statistically significant at the 1% level.

For the main specification, the average growth rate of 0.235 corresponds to an increase (in levels) of similarity of 0.125. The observed difference in levels is 0.105.

We also run a similar exercise to look at heterogeneity in the convergence of deliberative speech to the public summaries by member characteristic. Specifically, we compare the public summaries to the the speech of the chair, non-chairs, and those with above or below median experience on the FOMC. We do not find substantively different patterns across these groups, lending support to the argument that convergence in outcomes at

**Figure 3.3:** Approximate and observed growth for cosine and generalized cosine similarity



*Notes:* We compare the similarity of the public and private documents in each meeting, relative to the similarity of the benchmark averages. The  $y$ -axis is the true growth of similarity, and the  $x$ -axis is the first-order approximation to growth, following Equation 3.3. The left panel is the unweighted cosine similarity, and the right panel uses the generalized cosine similarity, which takes into account semantic relations. The correlation between the approximation and truth is close to 0.9. The dashed line is the 45-degree line, and the solid line is a locally weighted scatterplot smoothing line.

the FOMC may be a consequence of the common shock.<sup>27</sup>

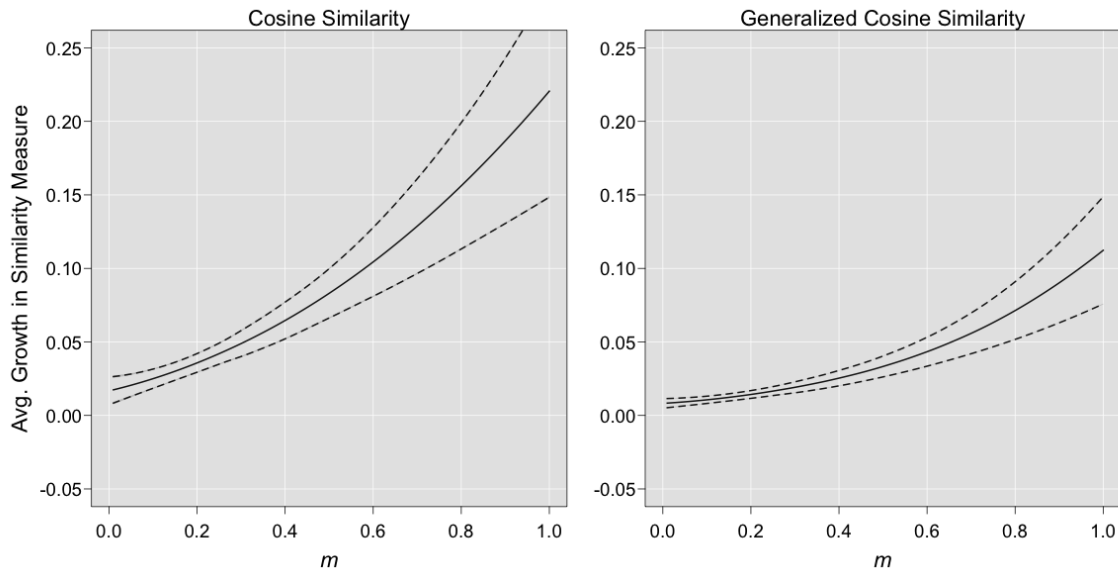
In Figure 3.3, we compare the approximate growth in cosine and generalized cosine similarity, using Equation 3.3, against the true growth in the respective measure. The correlation between the approximation and truth is close to 0.9. The approximation is very close to the truth and we are therefore comfortable analyzing its decomposition.

We also explore sensitivity to the weight placed on more-common words. In Figure 3.4, we explore how the treatment effect of the main specification varies as we put different weights on the words (decreasing  $m$  puts more weight on sparse terms). The solid line represents the coefficient on post-reform, and the dashed lines represent the 95%

<sup>27</sup>We do not report these results since they are not the focus of the paper, but are happy to provide them by request.

confidence interval. Increasing  $m$  is associated with increased similarity, suggesting that the “treatment” effect is not driven by changes among the rarely-used words.

**Figure 3.4:** *Extended cosine similarity of FOMC transcripts and corresponding public summaries from 1976–2007*



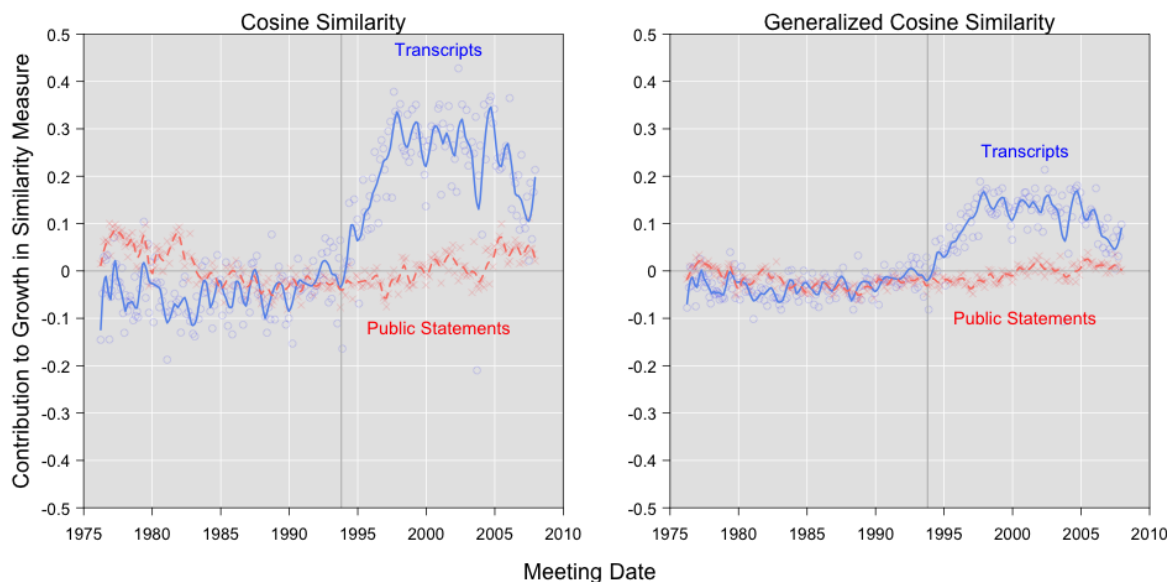
*Notes:* This figure plots the “treatment effect” of transparency under different weighting schemes. We calculate  $CS(p^m, q^m)$  for various levels of  $m$ , where  $m = \frac{1}{2}$  corresponds to the Atkinson Index and  $m = 1$  is Cosine Similarity. The  $x$ -axis corresponds to  $m$ , and the  $y$ -axis is the change in average similarity before and after the policy change. The dashed lines correspond to 95% confidence intervals.

### Decomposition by Document Type

In Figure 3.5, we decompose the change in similarity at the document level. We consider how the relationship between the documents would have changed if only the transcripts (blue line) or public statements (red line) evolved (holding fixed the other document). Both within the unweighted economics words (left panel) and including the semantic similarity weights (right panel), it is clear that almost the entirety of the increase in similarity comes through changes in the transcripts, and not changes in the public statements. With

generalized cosine similarity the result is qualitatively similar, although somewhat scaled down (given the additional information about cross-word relations).

**Figure 3.5:** *Decomposition of similarity growth into public and private document contributions*



*Notes:* This figure plots the first-order approximation to the Cosine (left panel) and Generalized Cosine (right panel) similarity of the public and private documents over time. The blue line corresponds to the transcripts' contribution, and the dashed red line to the public statements'. The vertical line corresponds to the policy change. Both panels suggest nearly all of the observed change in similarity is attributable to variation in the meeting transcripts after the transparency reform.

Another way to generate inference on our results is through permutation tests, which relax assumptions about the form of the sampling distributions of our test statistics. For each permutation, we preserve the actual distribution of word changes and gaps for each document, but randomly match them together.<sup>28</sup> Each gray dot in the left panels of Figure 3.6 represents the counterfactual change in similarity relative to the true baseline. The black dots represent the true change. Since the gaps are constant in the whole range (with the exception of the leave-out baselines for the 50 meetings right before the policy shock), this is a way to visualize the distribution of word-growths over time. It is clear

<sup>28</sup>That is, each  $n_{pt}^j$  is randomly matched to  $w_{pt}^i$ .

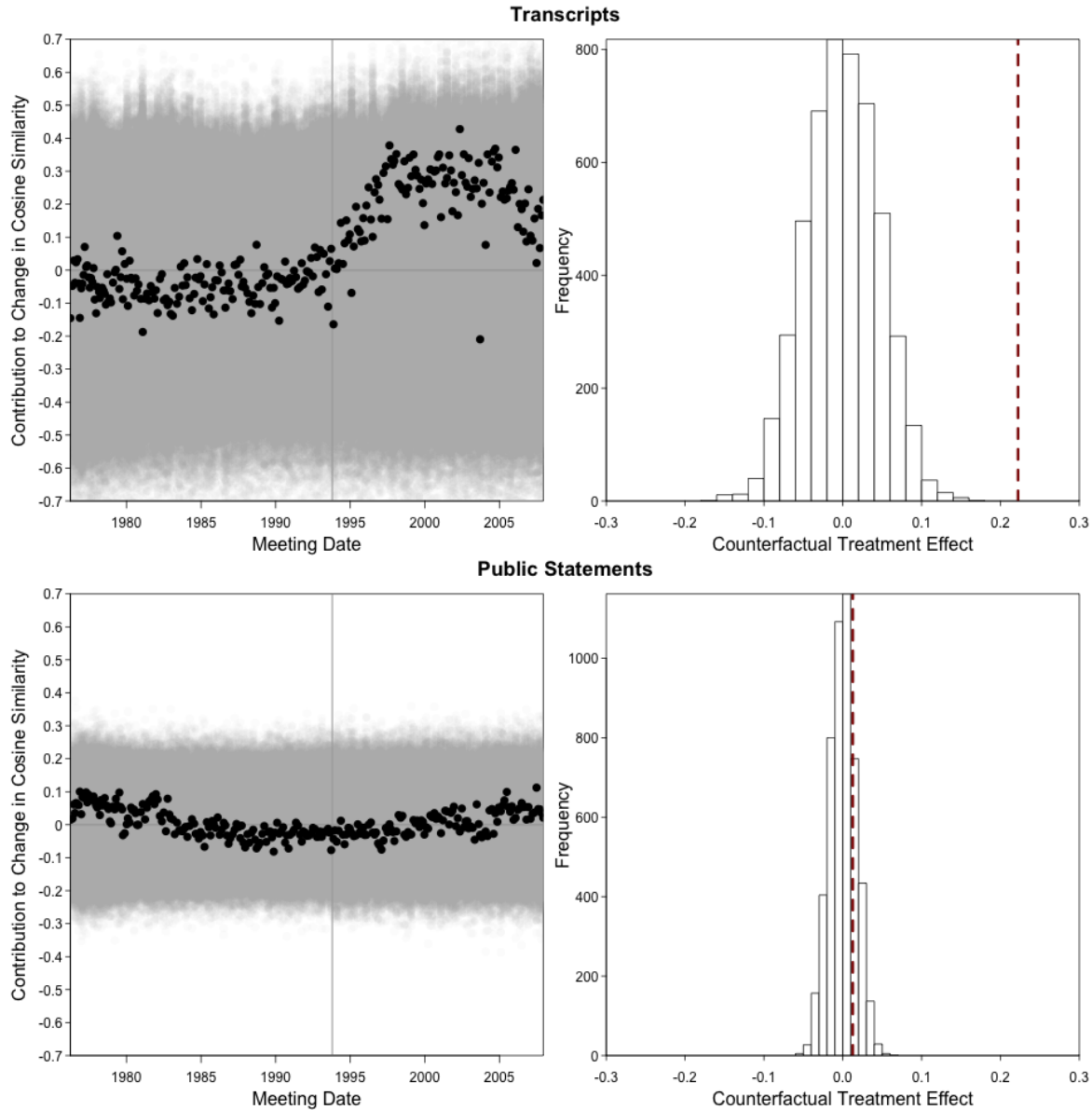
that the distribution is stable over time; the sets of potential growth over time looks much like a rectangle. The right panels show a histogram of the counterfactual “average treatment effect” for all of the permutations. Even though the range for any particular meeting is much larger in magnitude than the true observed change, in 5000 iterations the counterfactual overall average effect in the transcripts is never as large as the true treatment effect. In the public statements, 21.6% of the permutations are larger than the true effect for the cosine similarity, and 13.9% are larger using generalized cosine similarity. Note that, analytically, the expected permuted treatment effect is zero, since the expected value of the covariance between permuted word changes and gaps is zero.

### **Decomposition by Word**

Our growth decomposition allows us to specifically identify the most important words to the observed change in similarity. Figure 3.8 shows that most of the total change in similarity comes from a relatively small number of words, where for each word we subtract each word’s average contribution to growth in the pre-period from that from the post-period. For the transcripts’ contribution to similarity, roughly one hundred words are responsible for 90 percent of the total negative change in similarity, and roughly one hundred words are responsible for 90 percent of the total positive change in similarity, as indicated by the vertical lines. The positive change in similarity is about five times greater than the negative change, leading to the net change of approximately 0.2. Relative to the transcripts, the net change in similarity contributed by the public documents is negligible.

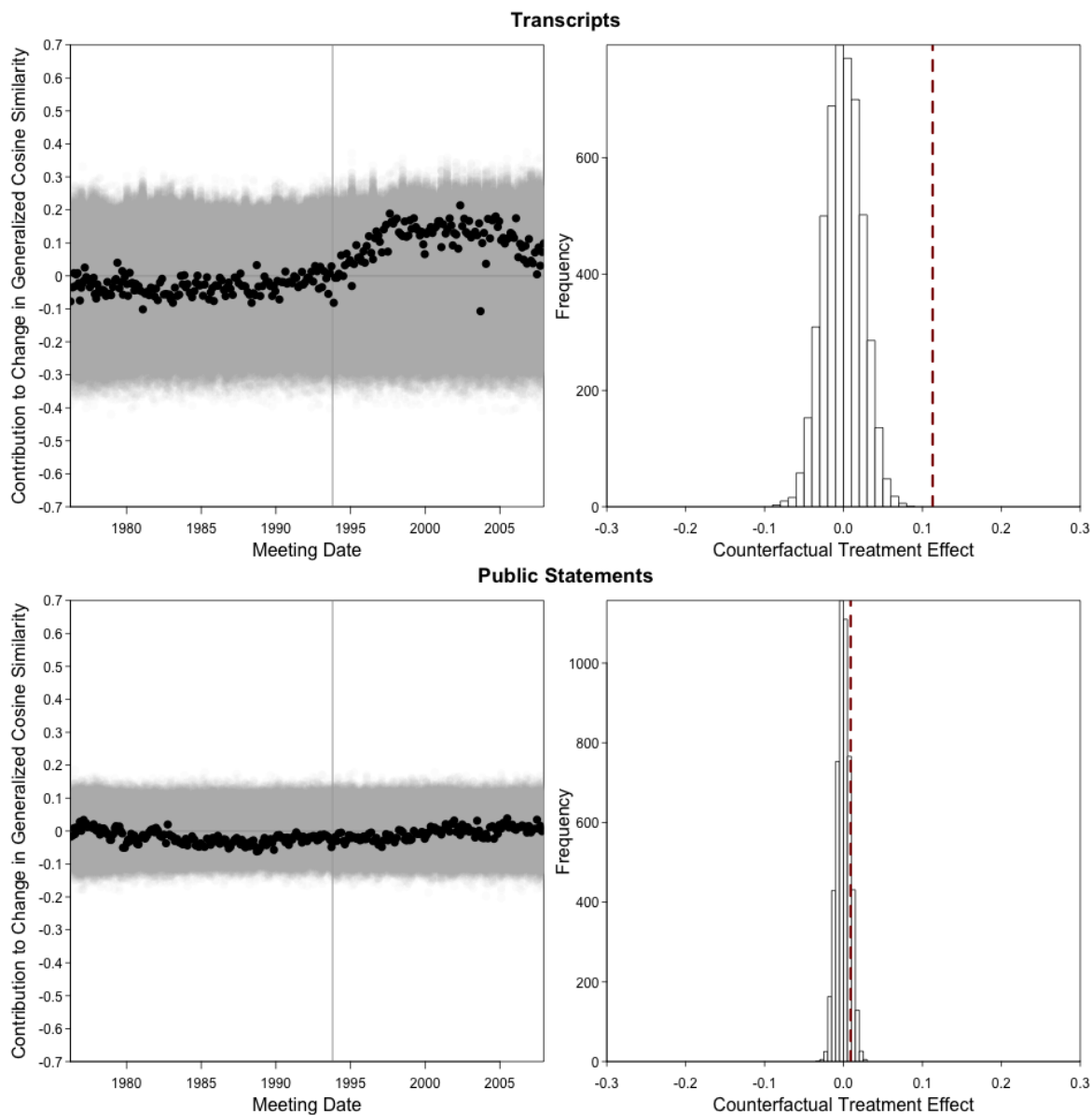
In Figure 3.9, we show the 10 words with the greatest positive and negative contributions to the growth in similarity in the transcripts and public statements, calculated using Equation 3.1. We also distinguish words whose use declined from those whose use increased. The top panel shows the results using the cosine similarity measure, the bottom panel adds in the weights for the generalized cosine similarity measure. As shown in the previous figures in this section, it is clear that the majority of the change is

**Figure 3.6:** *Permutation tests of the change in cosine similarity*



*Notes:* This figure plots, on the left, the permuted counterfactual change in similarity for the transcripts (top panel) and public statements (lower panel). Each light gray dot corresponds to one meeting's placebo change in similarity for one of the 5000 permutations. The black dots represent the observed values in the sample. The figures on the right show the distribution of permuted "treatment effects," comparing the average difference in similarity for the fifty meetings immediately before the policy change to the fifty meetings immediately after. The vertical, dotted-line line represents the observed effect in the data.

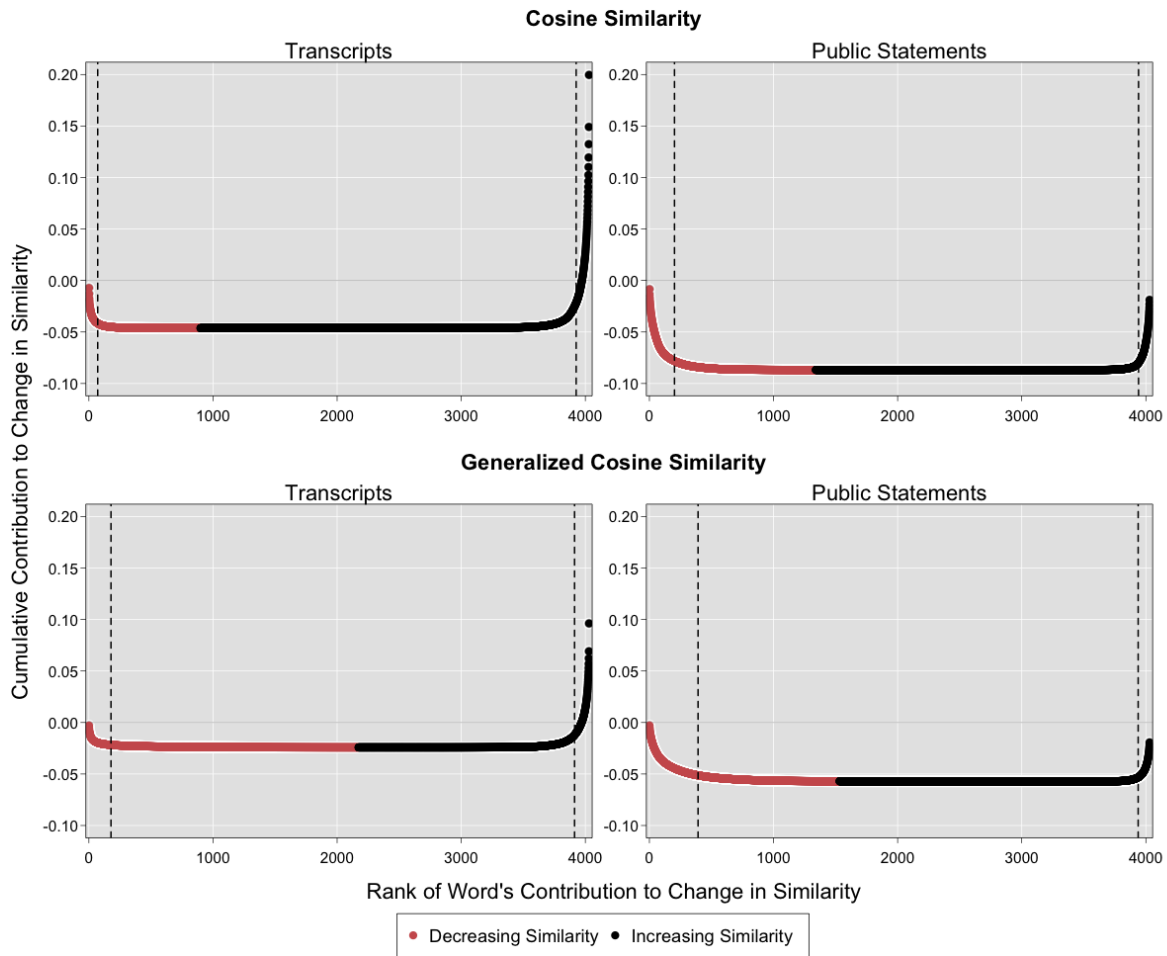
**Figure 3.7:** *Permutation tests of the change in generalized cosine similarity*



*Notes:* This figure plots, on the left, the permuted counterfactual change in similarity for the transcripts (top panel) and public statements (lower panel). Each light gray dot corresponds to one meeting’s placebo change in similarity for one of the 5000 permutations. The black dots represent the observed values. The figures on the right show the distribution of permuted “treatment effects,” comparing the average difference in similarity for the fifty meetings immediately before the policy change to the fifty meetings immediately after. The vertical line represents the true effect in the data.



**Figure 3.8:** Cumulative contribution of word's similarity growth



*Notes:* This figure shows the cumulative change in average similarity. Words (in the transcripts on the left, and public statements on the right) are ordered on the  $x$ -axis according to their contribution to the aggregate change in similarity of the public and private documents. On the  $y$ -axis is the cumulative sum of similarity until that rank. The dashed lines represent 90% of the total decrease (left) and increase (right) in similarity.

coming from the transcripts. Furthermore, there are a few words which are substantially responsible for the change. The words responsible for positive changes in similarity and whose usage increased were mostly related to economics, such as “growth,” “market,” and “price” (in other words, before the reform “growth” had been underused in the transcripts relative to the public statements). The words responsible for positive changes in similarity but whose usage declined were mostly not related to economics, such as “think,” “say,” “that,” and “don’t.” Our use of the generalized cosine similarity measure (that accounts for cross-word similarities) does not meaningfully change which words are the major contributors to similarity growth over time.<sup>29</sup>

The decline in the word “think” is consistent with anecdotal evidence that the FOMC meetings became less of a conversation after transparency, with members bringing in speeches of their own. The increase in common economics words suggests that the type of language that FOMC members decided to prepare was much more in line with the public statements along certain identifiable dimensions.

### 3.5 Conclusion

We develop two novel textual methods to examine how the sudden enforcement of the Sunshine Act in 1993 affected communication in FOMC meetings. Our goal is to identify a few words or dimensions that were most responsible for the change. We develop a theoretical model to predict how FOMC members would respond to the reform, generates a rationale for using cosine-similarity to measure the effects of the policy change. Instead of first grouping the language into a few clusters, and then focusing our analysis on those clusters (as is increasingly popular in computational linguistics research, e.g., Hansen, McMahan and Prat (2014)), this method allows us to group the data by how much it

---

<sup>29</sup>“Think” shows up in the ODE twice: first in the definition of “Club of Rome” (which is a think tank), and “bad debt” (whose definition includes a discussion of what creditors might think).

**Figure 3.9:** Relative contributions by individual words to growth in cosine similarity and generalized cosine similarity



*Notes:* This figure plots individual word/document dyad's contribution to the average change in Cosine (top) and Generalized Cosine (bottom) Similarity after the policy change. The 10 words that led to the largest increase and decrease in similarity are represented. The words with a corresponding up (down) arrow are those whose usage increased (decreased) in that document.

was affected by the policy change. We find that a few words were primarily responsible for the changed behavior after transparency reform. We develop methods that allow researchers to adjust for semantic similarity across words, which we implement using the definitions in the Oxford Dictionary of Economics. Accounting for semantic similarity does not qualitatively change our results on the effects of the policy reform.

In particular, we find that transparency led the previously private FOMC meeting conversations to become more similar to the always-released public statements. In our setting, we found that the proportion of speech related to economics increased after the policy change for both the chair and the non-chairs. To uncover the dimensions of this change, we decomposed the change in cosine similarity of the public and private documents into word level contributions. We found that most of the change in behavior came from FOMC members shifting their speech towards popular economic topics, such as “inflation” and “growth,” and away from hedging language such as “think.” These results are robust to restricting our analysis to terms in the Oxford Dictionary of Economics, and to allowing our similarity metric to account for cross-word similarities with a relationship weight matrix. Our proposed methods extend and add robustness to any analysis considering the similarity of agents over time.

## Appendices

### A Appendix to Chapter 1

#### A.1 Supplementary Analyses

##### Scaling Firms by their Product Portfolios

Given the analysis in the previous subsections, a natural question is to ask is to what degree individual products sold by firms map to estimated firm-level causal effects. Put another way, are sales of individual products “predictive” of abnormal financial returns across our event studies? To inspect this question, we estimate linear equations of the form:

$$\begin{aligned}\hat{\phi}_i &= \beta_0 + \sum_k \beta_k \left( \sum_{q \in Q} \tilde{d}_{i,q}^k \right) + u_i \\ \tilde{\phi}_i &= \beta_0 + \sum_k \left( \beta_k \cdot \tilde{d}_i^k \right) + u_i,\end{aligned}\quad (4)$$

where  $\tilde{\phi}_i = (\hat{\phi}_i - \bar{\phi})/s_\phi$ ,  $\bar{\phi} = \frac{1}{n-1} \sum_i \hat{\phi}_i$ ,  $s_\phi = \sqrt{\frac{1}{n-1} \sum_i (\hat{\phi}_i - \bar{\phi})^2}$ ,  $\tilde{d}_i^k = (d_i^k - \bar{d}^k)/s_{d_k}$ , and  $s_{d_k} = \sqrt{\frac{1}{n} \sum_i (d_i^k - \bar{d}^k)^2}$ . In words, the estimated coefficients  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$  from Equation 4 are in terms of standardized units, as the outcomes and firm-level product totals are rescaled according to their means and sample standard deviations. For this analysis, as in Section 1.4.2, we limit our attention to firms with pre-event market capitalizations over 100 Million USD and have been publicly traded for at least the prior 8 financial

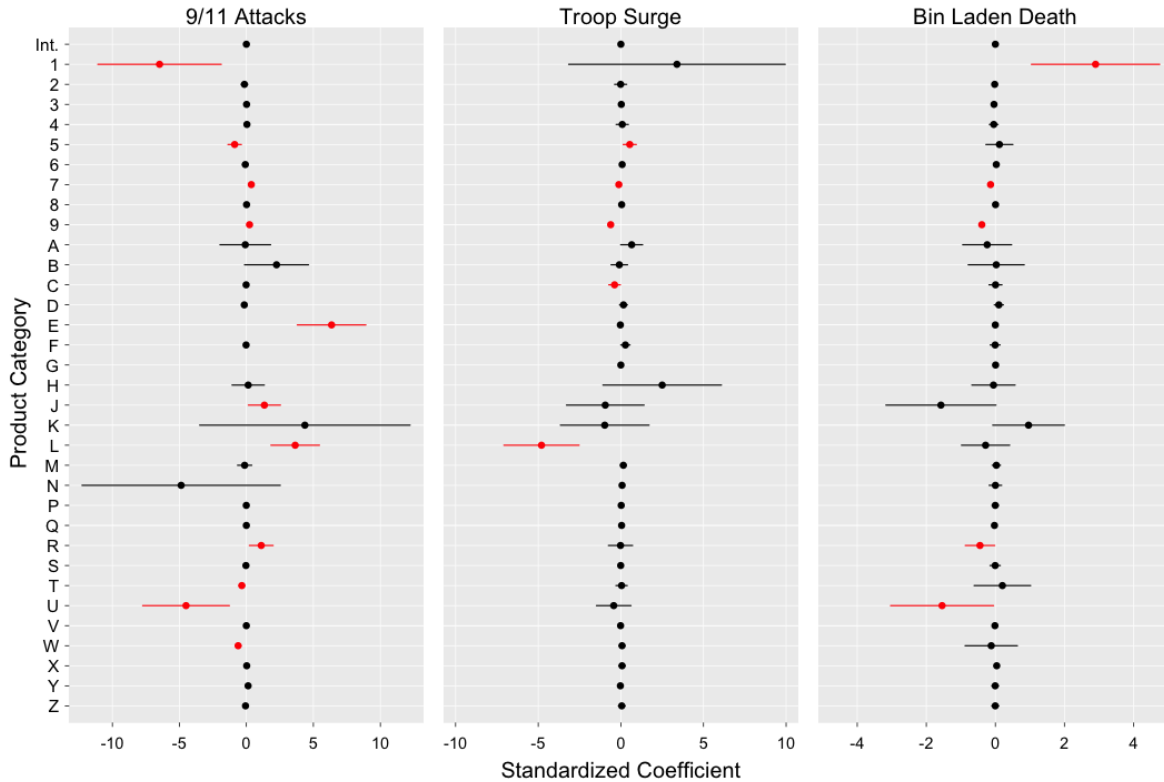
quarters. We impose the additional constraint that only firms with above-median revenue dependence measures  $\mathcal{R}_i$  are included in this analysis.

Results of this analysis are presented in Figure A.10. The vertical axis in each subplot indicates the single-digit product category coefficient. Across event studies, a subset of product categories appear to be consistently correlated with firm-level causal effects. Although such results should be taken with a grain of salt—as individual firms vary markedly in terms of their overall size, the products they produce, and their revenue dependence to the DoD—it is noteworthy that firm-level product totals are in the aggregate associated with effect estimates. Additional research is required to more precisely decouple the effects of individual products, but we leave that task for future study.

### **Procurement Competitiveness by Firm**

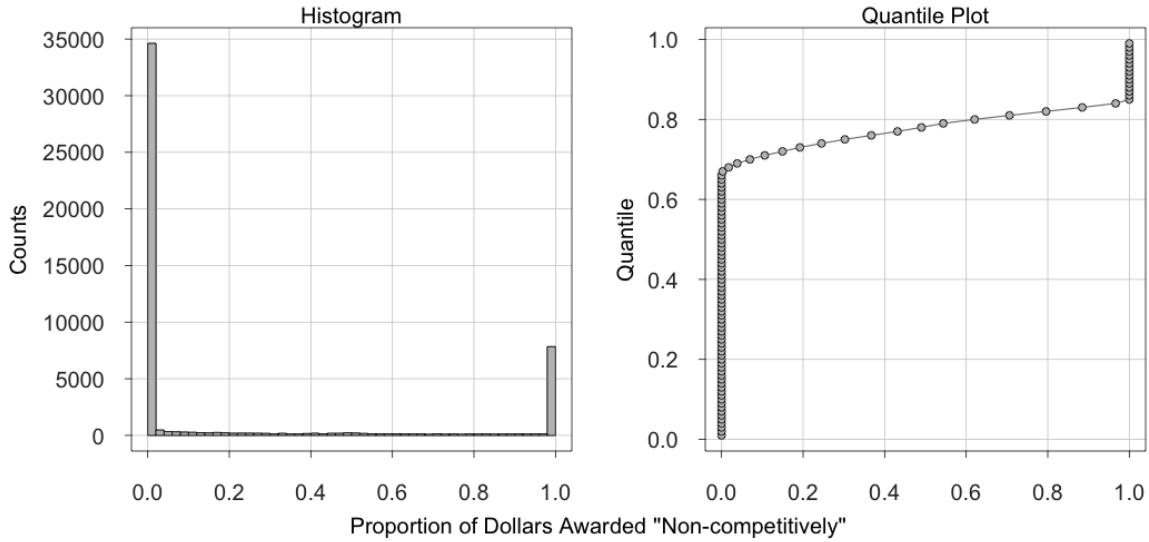
Between the start of FY2000 to September 10, 2001, the U.S. Department of Defense negotiated 668,418 contract actions with a cumulative nominal valuation of over \$278 Billion USD. Roughly 39% of these dollars were awarded through “non-competitive” procurement awards, and more than 50,000 unique firms received obligations over this interval. The figure below demonstrates that a typical firm in the sample receives either all or none of its contracts competitively (due to the fact that the modal firm in the FPDS data receives very few contracts).

**Figure A.10:** How Firm-level Product Sales Map to Causal Effects



Notes: Each dot in each subplot indicates a regression coefficient estimate for a given product category. **Highlighted** estimates indicate a regression coefficient is statistically significant, with 95% confidence intervals surrounding each estimate. Product categories are aggregated to single-digit PSCs for this analysis. See Table A.1 for additional information on the (two-digit) product categories contained in the data. Across event studies, a subset of product categories appear to be predictive of individual causal effects. For example, Category 1 refers to products that may be considered as offensive products (e.g., weapons, air ships, missiles). Category U refers to education and training services. Category 5 refers to raw materials needed in construction services (e.g., wood, tools, electrical components). Category 7 refers to food and indoor services.

**Figure A.11:** *Proportion of Each Firm's Total Obligations Awarded "Non-competitively" (FY2000 to September 10, 2001)*

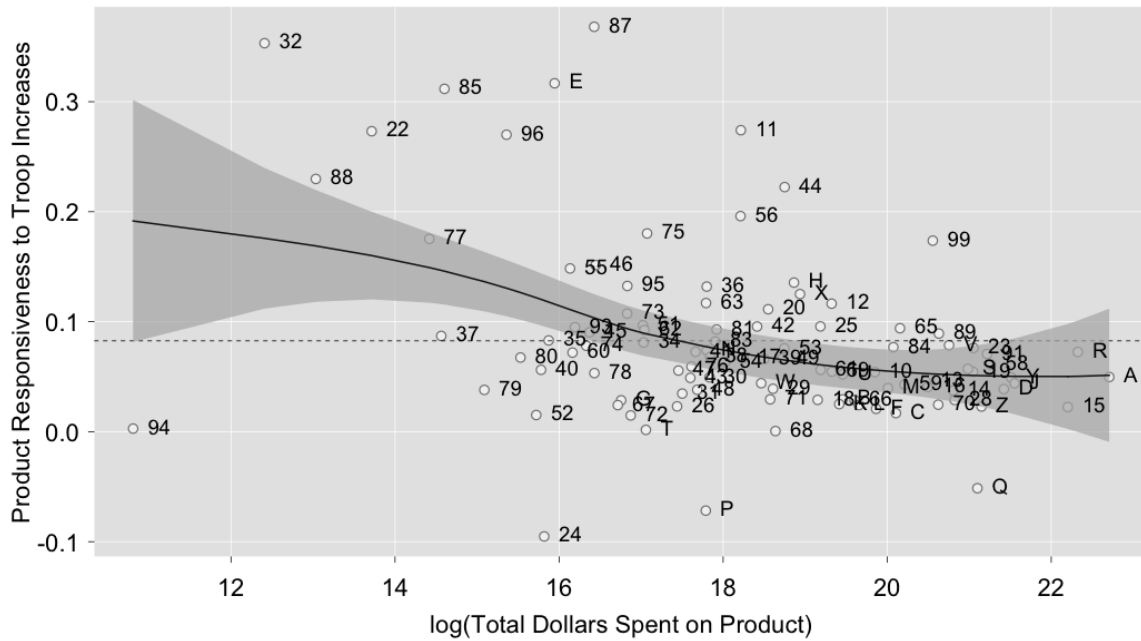


*Notes:* Utilizing all DoD contract records, firms are scaled by the proportion of all their contract dollars that were awarded non-competitively. Equivalent patterns exist across fiscal years in the FPDS data.



## A.2 Product Responsiveness to War Escalation

**Figure A.12:** How Product Purchases Grow Alongside Troop Levels (All DoD Obligations Prior to Iraq Surge Announcement)



*Notes:* Each dot represents a two-digit product category, as outlined in Table A.1. Tabulating all DoD obligations to individual product categories at the monthly level, this figure plots the results of a log-log regression against the total number of troops in Afghanistan and Iraq in a given month:  $\log(d_{kt} + 1) = b_{0k} + b_{1k} \cdot \log(\text{Troops}_t + 1) + u_{kt}$ . The coefficient  $b_{1k}$  represents an estimate of how a percent increase in total troop levels is expected to map to growth in product expenditures. Overall, increasing troop levels by 100% is expected to increase outlays to a typical product code by roughly 8%.

## A.3 Data Sources and Cleaning

### Records of Defense Products and Services (PSCs)

The core of the product-level data used in this analysis are taken from the Federal Procurement Data System (FPDS), the official source of U.S. government procurement data for all contracts signed with Federal departments and agencies, as mandated in 41 U.S.C. 401.<sup>30</sup> For this analysis, all listed contract actions for each fiscal year between FY2000 and FY2014 were collected and cleaned into a single database. For each stated fiscal year, listed procurements begin on October 1 of the previous year and span to September 30 of the same year (e.g., FY2000 spans from October 1, 1999 to September 30, 2000).

Entries in this procurement dataset (i.e., “the rows” of the database) are recorded at the contract-action level. In an average fiscal year in this sample there are several million contract actions recorded in the procurement data, with the fewest number of actions occurring near the start of the data and the greatest number during the height of the Iraq War. For each contract action, over 200 characteristics of the event are recorded in the FPDS, which include among other details: the date the contract was signed, the contract serial number (i.e., a unique award ID for each action), the recipient firm’s name, the parent company of the recipient firm, the location of the contract’s action, the nominal value of the dollars obligated (in U.S. dollars), the funding agency within the DoD (e.g., Army, Navy, Air Force, DARPA), whether the contract was awarded through competitive bids (yes or no), in addition to the Product or Service Code (PSC) of the awarded contracts. The PSC is the official government label of the product or service procured in a given purchase or award, and PSC codes are standardized across federal agencies. PSCs are recorded at the two-digit and four-digit level, the latter of which is a

---

<sup>30</sup>This resource is accessible online at: “Federal Procurement Data System - Next Generation,” [https://www.fpds.gov/fpdsng\\_cms/index.php/en/](https://www.fpds.gov/fpdsng_cms/index.php/en/). Full documentation of 41 U.S.C. 401 can be found in <http://www.gpo.gov/fdsys/granule/USCODE-1997-title41/USCODE-1997-title41-chap7-sec401>.

more granular measure of the individual product or service provided. For example, the four-digit PSC “1005” refers to “Guns, through 30mm”, which refers to a class of firearms that include “Machine guns; Brushes, Machine Gun and Pistol.” Similarly, the four-digit PSC “9130” refers to “Liquid Propellants and Fuels, Petroleum Base”, which includes “All Aviation Gasoline; JP-1, 3, 4, and 5 Jet Fuel; Combat Vehicle and Automotive Gasoline (all types and grades); Liquid Propellants, Bulk; Liquid Propellants, predetermined to specify quantity and quality, packaged in reusable containers.” Table A.1 provides a summary of all two-digit PSCs that appear in the study sample. Figure A.13 provides a visual summary of how DoD obligations move to individual PSCs over time.

The PSC-level information, in conjunction with information on the recipient firms of individual awards, are especially fruitful in this analysis. The PSC-level data (i.e., product-code specific outlays to a firm) is used to scale the relations between firms across the defense industry. In the data, over four thousand unique PSC character strings appear at least once in the fifteen years of procurement data; some of these character strings are redundant since reporting rules and naming conventions for these PSC have changed over time, as outlined in the “Reference Data Source Information” provided by the FPDS.<sup>31</sup> For example, the PSC entitled “1550: DRONES” (which is defined as “Piloted aircraft and guided missiles converted to drone use”) was renamed “1550: UNMANNED AIRCRAFT” on October 1, 2011. With renaming conventions being observable, PSC codes are cleaned and standardized to allow for consistent measures of firm-product outlays over time. Once PSCs have been adjusted for differences in naming conventions, roughly 3,000 unique PSCs are left for analysis. A graph showing aggregate, PSC-level daily outlays made by the DoD can be found in Figure A.13.

---

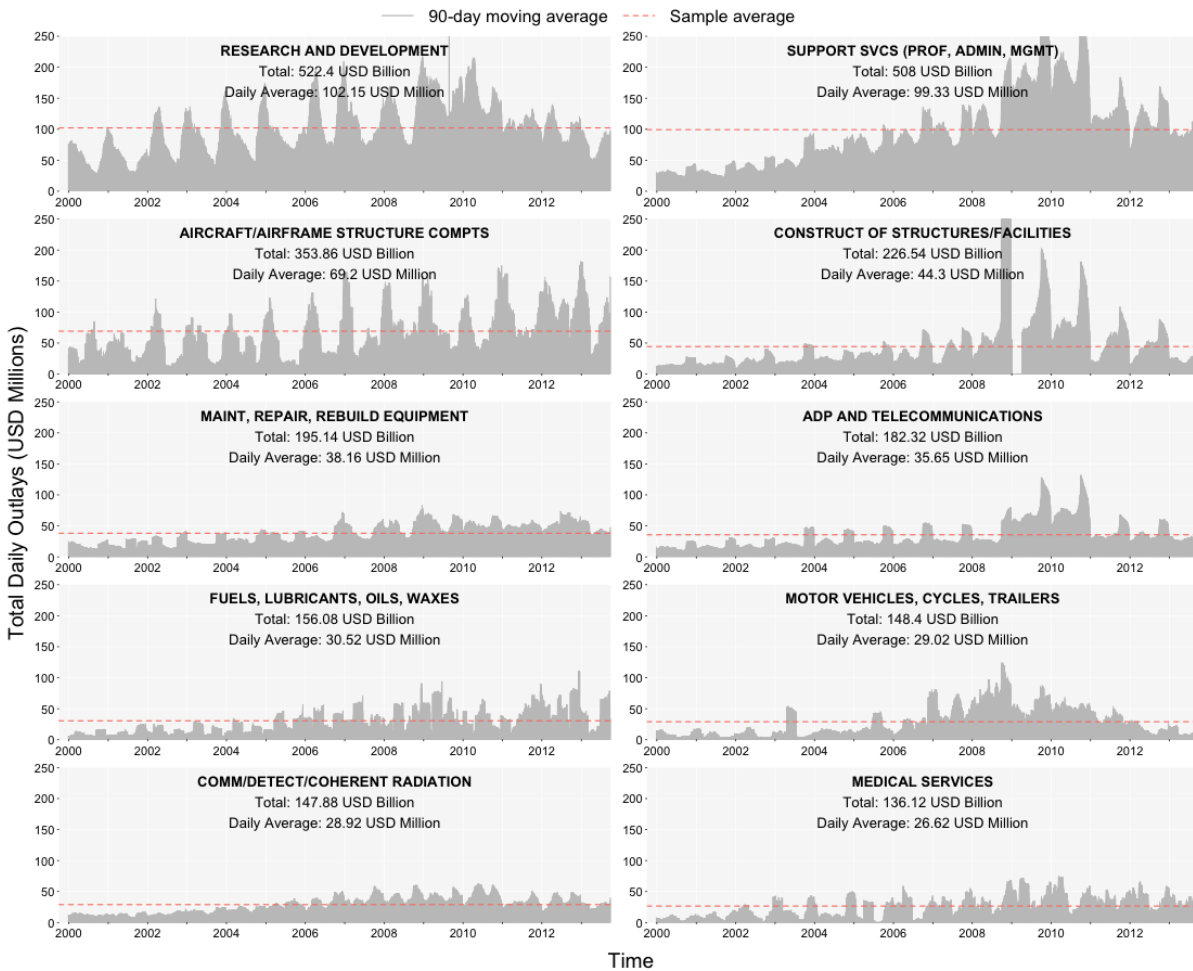
<sup>31</sup>See, for example, Data Element Number 8A available online at [https://www.fpds.gov/wiki/index.php/PSC,\\_NAICS\\_and\\_more](https://www.fpds.gov/wiki/index.php/PSC,_NAICS_and_more).

**Table A.1: Two-Digit Product Codes in DoD Procurement Data**

PSC	Meaning	PSC	Meaning
10	WEAPONS	68	CHEMICALS AND CHEMICAL PRODUCTS
11	NUCLEAR ORDNANCE	69	TRAINING AIDS AND DEVICES
12	FIRE CONTROL EQPT.	70	ADP EQPT/SOFTWARE/SUPPLIES AND EQPT
13	AMMUNITION AND EXPLOSIVES	71	FURNITURE
14	GUIDED MISSLES	72	HOUSEHOLD/COMMERC FURNISH/APPLIANCE
15	AIRCRAFT/AIRFRAME STRUCTURE COMPTS	73	FOOD PREPARATION/SERVING EQPT
16	AIRCRAFT COMPONENTS/ACCESSORIES	74	OFFICE MACH/TEXT PROCESS/VISIB REC
17	AIRCRAFT LAUNCH/LAND/GROUND HANDLE	75	OFFICE SUPPLIES AND DEVICES
18	SPACE VEHICLES	76	BOOKS, MAPS, OTHER PUBLICATIONS
19	SHIPS, SMALL CRAFT, PONTOON, DOCKS	77	MUSICAL INST/PHONOGRAPH/HOME RADIO
20	SHIP AND MARINE EQUIPMENT	78	RECREATIONAL/ATHLETIC EQPT
22	RAILWAY EQUIPMENT	79	CLEANING EQPT AND SUPPLIES
23	MOTOR VEHICLES, CYCLES, TRAILERS	80	BRUSHES, PAINTS, SEALERS, ADHESIVES
24	TRACTORS	81	CONTAINERS/PACKAGING/PACKING SUPPL
25	VEHICULAR EQUIPMENT COMPONENTS	83	TEXTILE/LEATHER/FUR; TENT; FLAG
26	TIRES AND TUBES	84	CLOTHING/INDIVIDUAL EQPT, INSIGNIA
28	ENGINES AND TURBINES AND COMPONENT	85	TOILETRIES
29	ENGINE ACCESSORIES	87	AGRICULTURAL SUPPLIES
30	MECHANICAL POWER TRANSMISSION EQPT	88	LIVE ANIMALS
31	BEARINGS	89	SUBSISTENCE
32	WOODWORKING MACHINERY AND EQPT	91	FUELS, LUBRICANTS, OILS, WAXES
34	METALWORKING MACHINERY	93	NONMETALLIC FABRICATED MATERIALS
35	SERVICE AND TRADE EQPT	94	NONMETALLIC CRUDE MATERIALS
36	SPECIAL INDUSTRY MACHINERY	95	METAL BARS, SHEETS, SHAPES
37	AGRICULTURAL MACHINERY AND EQPT	96	ORES, MINERALS AND PRIMARY PRODUCTS
38	CONSTRUCT/MINE/EXCAVATE/HIGHWAY EQPT	99	MISCELLANEOUS
39	MATERIALS HANDLING EQPT	A	RESEARCH AND DEVELOPMENT
40	ROPE, CABLE, CHAIN, FITTINGS	B	SPECIAL STUDIES/ANALYSIS, NOT RandD
41	REFRIG, AIR CONDIT/CIRCULAT EQPT	C	ARCHITECT/ENGINEER SERVICES
42	FIRE/RESCUE/SAFETY: ENVIRO PROTECT	D	ADP AND TELECOMMUNICATIONS
43	PUMPS AND COMPRESSORS	E	PURCHASE OF STRUCTURES/FACILITIES
44	FURNACE/STEAM/DRYING; NUCL REACTOR	F	NATURAL RESOURCES MANAGEMENT
45	PLUMBING, HEATING, WASTE DISPOSAL	G	SOCIAL SERVICES
46	WATER PURIFICATION/SEWAGE TREATMENT	H	QUALITY CONTROL, TEST, INSPECTION
47	PIPE, TUBING, HOSE, AND FITTINGS	J	MAINT, REPAIR, REBUILD EQUIPMENT
48	VALVES	K	MODIFICATION OF EQUIPMENT
49	MAINT/REPAIR SHOP EQPT	L	TECHNICAL REPRESENTATIVE SVCS.
51	HAND TOOLS	M	OPERATION OF GOVT OWNED FACILITY
52	MEASURING TOOLS	N	INSTALLATION OF EQUIPMENT
53	HARDWARE AND ABRASIVES	P	SALVAGE SERVICES
54	PREFAB STRUCTURES/SCAFFOLDING	Q	MEDICAL SERVICES
55	LUMBER, MILLWORK, PLYWOOD, VENEER	R	SUPPORT SVCS (PROF, ADMIN, MGMT)
56	CONSTRUCTION AND BUILDING MATERIAL	S	UTILITIES AND HOUSEKEEPING
58	COMM/DETECT/COHERENT RADIATION	T	PHOTO, MAP, PRINT, PUBLICATION
59	ELECTRICAL/ELECTRONIC EQPT COMPNTS	U	EDUCATION AND TRAINING
60	FIBER OPTIC	V	TRANSPORT, TRAVEL, RELOCATION
61	ELECTRIC WIRE, POWER DISTRIB EQPT	V1	TRANSPORTATION OF THINGS
62	LIGHTING FIXTURES, LAMPS	W	LEASE/RENT EQUIPMENT
63	ALARM, SIGNAL, SECURITY DETECTION	X	LEASE/RENT FACILITIES
65	MEDICAL/DENTAL/VETERINARY EQPT/SUPP	Y	CONSTRUCT OF STRUCTURES/FACILITIES
66	INSTRUMENTS AND LABORATORY EQPT	Z	MAINT, REPAIR, ALTER REAL PROPERTY
67	PHOTOGRAPHIC EQPT		

Notes: Two-digit PSCs are presented in this table to provide a visual summary of the goods sold to the U.S. government. Analysis in the main body of the text utilizes four-digit PSCs, however. The PSC codes and descriptions are taken as presented in original Federal Procurement Data System (FPDS) summary files.

**Figure A.13: Daily Outlays to Top-Grossing PSCs (FY2000 onwards)**



*Notes:* This figure plots the daily DoD outlays to the top-grossing product service codes (PSCs) between 2000 and the end of 2013. For presentational purposes, only the top 10 PSCs are presented in this figure. In general, there are noticeable increases in both the frequency and average daily outlays to top PSCs following the start of the Afghanistan and Iraq Wars, although the degree of growth varies by product.

## Obtaining Firm-Level Procurement Records

Although FPDS procurement data are collected for a wide variety of firms (that include, for example, both publicly traded and private firms), more fine-grain financial characteristics, by necessity, can only be collected for the subset of firms that are publicly traded. The main specification in the statistical analysis involves a sample of public and private firms together.<sup>32</sup> Some care was required to match the unique set character strings representing FPDS company names (i.e., as it is noted in the raw data, the “ParentRecipientOrCompany Name” for a given contract action) with their associated company label in the lobbying data (which is described in the next subsection). Because the FPDS indexes firm contracts with proprietary company codes (assigned by the business information company, Dun & Bradstreet), there is no straightforward way to merge FPDS data with other official U.S. government documents (e.g., using a CIK key from the SEC) or financial market records.<sup>33</sup> Details of the primary name matching algorithm can be found in the Appendix. The script was utilized to first identify the set of plausible company name matches, while finalized matches were determined through manual verification.

With PSC-specific outlay data measured at the time and company level, one is able to construct measures of DoD connectedness using both information on aggregate product-level obligations and company-level PSC outlays. For each public company in each year, one has measure of what share of its total revenues come from the DoD, and which products are most responsible for that dependency. A stylized representation of these dependencies can be found in Figure A.18 (in the Appendix), which shows the PSCs that contribute the most to each firm’s total DoD revenue. Overall it is quite common to see multi-product firms “specialize” in substantively-related products. For example,

---

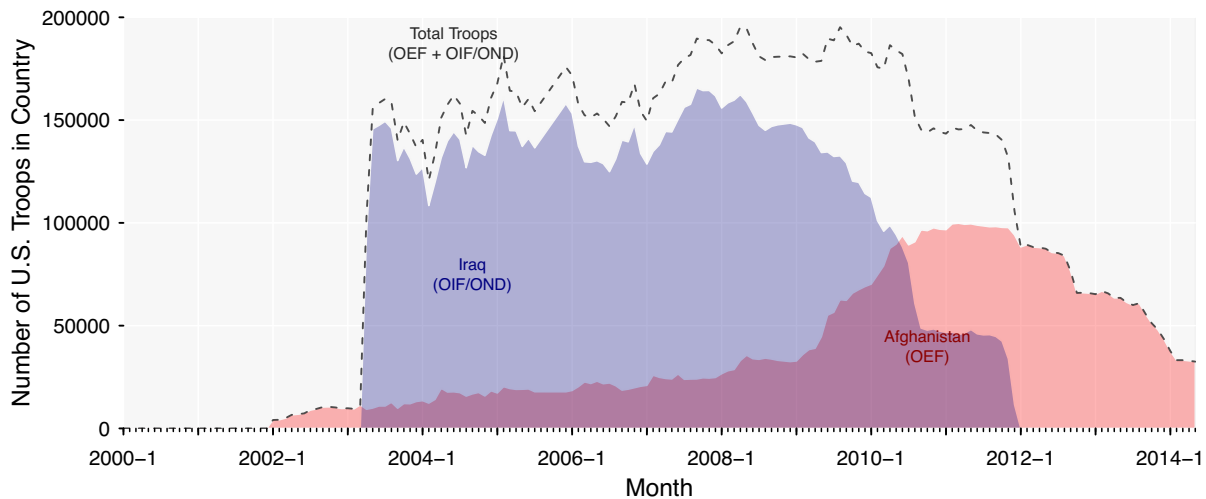
<sup>32</sup>However, more nuanced analyses or heterogenous effects and sensitivity analyses are possible for the set of public firms in the data. These analyses are a focus of continuing research.

<sup>33</sup>See, for example, Bloom, Schankerman and Van Reenen (2013) for a similar discussion on the challenge of cleaning and merging such data.

Lockheed Martin Corporation is a major producer of fixed-wing aircraft, space vehicles, and also sells maintenance and aircraft repair services, but it also sells hundreds of other products to the DoD. Firms like Oshkosh Corporation, by contrast, specialize in armored vehicles, and provide close to no other services to the DoD.

## War Data: U.S. Troop Levels in Afghanistan and Iraq

**Figure A.14: Monthly U.S. Troop Levels in Afghanistan and Iraq**



*Notes:* This figure plots the average number of U.S. troops deployed “in country” in the Afghanistan and Iraq wars over time, where the blue polygon represents troop levels in the Iraq War and the red polygon represents Afghanistan. The dotted line above both polygons represents the total troops in both countries. The figure highlights how Iraq received the majority of U.S. troop attention until midway through 2010, at which point Afghanistan had a greater number of troops deployed in country. These data are obtained from the Congressional Research Report “The Cost of Iraq, Afghanistan, and Other Global War on Terror Operations Since 9/11” (Belasco, 2014), which cites briefings made to Congress by the Joint Chiefs of Staff. The three-letter codes refer to the official mission names from the U.S. government: Operation Iraqi Freedom, Operation New Dawn, and Operation Enduring Freedom.

## Matching Firms Across Datasets

Figure A.15: R Script for Matching Firm Names Across Databases

```
1
2 # Define the fuzzy match function:
3 stem_list <- list(c(" COS$", " COMPANY$", " COMPS$"),c(" LLC$", " L L C$"),c(" INCS$", " INCORPORATED$", " Inc$"), c(" CORP$", "
  CORPORATIONS$"), c(" SERVICES$", " SERVICE$"),c(" LTD$", " LIMITED$"), c("^THE ", " THES$"),c(" PLC$"))
4
5 fuzzyNameMatch <- function(search_pattern, target_vector, stem_list=stem_list){
6   agrep_indices <- agrep(paste0("",search_pattern), target_vector)
7   agrep_matches <- target_vector[agrep_indices]
8   if (length(agrep_matches)>0){
9     agrep_match_dat <- data.frame(search_name=search_pattern, matched_name=agrep_matches, dist=as.vector(adist(gsub(" ", "", search_
      pattern), unlist(lapply(agrep_matches, function(y){gsub(" |[:punct:]" , "", y)[[1]]}))), index=agrep_indices)
10    agrep_match_dat <- arrange(agrep_match_dat, dist)
11    agrep_stem_fix_matches <- lapply(agrep_match_dat$matched_name, function(y){unique(sort(unlist(lapply(stem_list, function(x){gsub(
      paste0(x, collapse="|" , "", y))))))})
12    fix_matches <- unique(unlist(lapply(stem_list, function(x){gsub(paste0(x, collapse="|" , "", search_pattern)})))
13    is_agrep_stem_match <- unlist(lapply(agrep_stem_fix_matches, function(x){length(intersect(x, fix_matches))>0}))
14    is_agrep_stem_match_ignoreSpace <- unlist(lapply(lapply(agrep_stem_fix_matches, function(y){gsub(" |[:punct:]" , "", y)}),
      function(x){length(intersect(x, lapply(fix_matches, function(y){gsub(" |[:punct:]" , "", y)}))>0}))
15    stem_dist <- unlist(lapply(lapply(agrep_stem_fix_matches, function(y){gsub(" |[:punct:]" , "", y)}), function(x){min(adist(x,
      lapply(fix_matches, function(y){gsub(" |[:punct:]" , "", y)}))}))
16    match_dat_fuzzy <- data.frame(agrep_match_dat, stem_dist= stem_dist, is_stem_match=is_agrep_stem_match, is_stem_match_ignoreSpace=
      is_agrep_stem_match_ignoreSpace)
17    match_dat_fuzzy$is_grep_match <- grepl(gsub(" |[:punct:]" , "", search_pattern), lapply(match_dat_fuzzy$matched_name, function(x){
      gsub(" |[:punct:]" , "", x)}))
18    match_dat_fuzzy <- arrange(match_dat_fuzzy, is_grep_match, decreasing=T)
19    match_dat_fuzzy <- arrange(match_dat_fuzzy, is_stem_match_ignoreSpace, decreasing=T)
20    match_dat_fuzzy <- arrange(match_dat_fuzzy, is_stem_match, decreasing=T)
21    match_dat_fuzzy$search_name <- as.character(match_dat_fuzzy$search_name)
22    match_dat_fuzzy$matched_name <- as.character(match_dat_fuzzy$matched_name)
23    match_dat_fuzzy$is_exact_match <- (match_dat_fuzzy$search_name==match_dat_fuzzy$matched_name)
24    match_dat_fuzzy <- head(match_dat_fuzzy, 20)
25  } else {
26    match_dat_fuzzy <- data.frame(search_name= search_pattern, matched_name=NA, dist=NA, stem_dist=NA, index=NA, is_stem_match=NA , is
      _stem_match_ignoreSpace=NA, is_exact_match=NA, is_grep_match=NA)
27  }
28  match_dat_fuzzy
29 }
```



## A.4 Summary of Covariates Included in Final BSTS Models

**Table A.2:** *BSTS Variable Inclusion Probabilities Across Events and Models*

### *Model 1: Baseline BSTS*

Variable	9/11 Attacks				Troop Surge				Bin Laden Death			
	Mean	10%	50%	90%	Mean	10%	50%	90%	Mean	10%	50%	90%
Market <sub>t</sub>	0.05	0	0.01	0.09	0.40	0.02	0.17	1	0.69	0.04	0.99	1
PCA <sub>1</sub>	0.29	0.01	0.09	1	0.35	0.04	0.18	1	0.07	0.01	0.04	0.15
PCA <sub>2</sub>	0.44	0.02	0.25	1	0.24	0.03	0.13	0.70	0.37	0.03	0.22	0.97
PCA <sub>3</sub>	0.22	0.01	0.06	0.98	0.27	0.01	0.09	1	0.10	0.02	0.05	0.25
PCA <sub>4</sub>	0.21	0.02	0.09	0.68	0.15	0.02	0.08	0.33	0.14	0.02	0.06	0.36
PCA <sub>5</sub>	0.12	0.01	0.04	0.32	0.19	0.01	0.07	0.70	0.12	0.01	0.05	0.32
PCA <sub>6</sub>	0.09	0.01	0.03	0.20	0.10	0.02	0.05	0.22	0.08	0.02	0.04	0.18
PCA <sub>7</sub>	0.21	0.01	0.06	0.79	0.19	0.01	0.05	0.80	0.06	0.01	0.03	0.11
PCA <sub>8</sub>	0.08	0.01	0.03	0.20	0.14	0.01	0.06	0.39	0.09	0.01	0.04	0.24
PCA <sub>9</sub>	0.15	0.01	0.04	0.46	0.20	0.01	0.07	0.73	0.05	0.01	0.02	0.07
PCA <sub>10</sub>	0.40	0.01	0.20	1	0.16	0.02	0.07	0.42	0.03	0	0.01	0.05

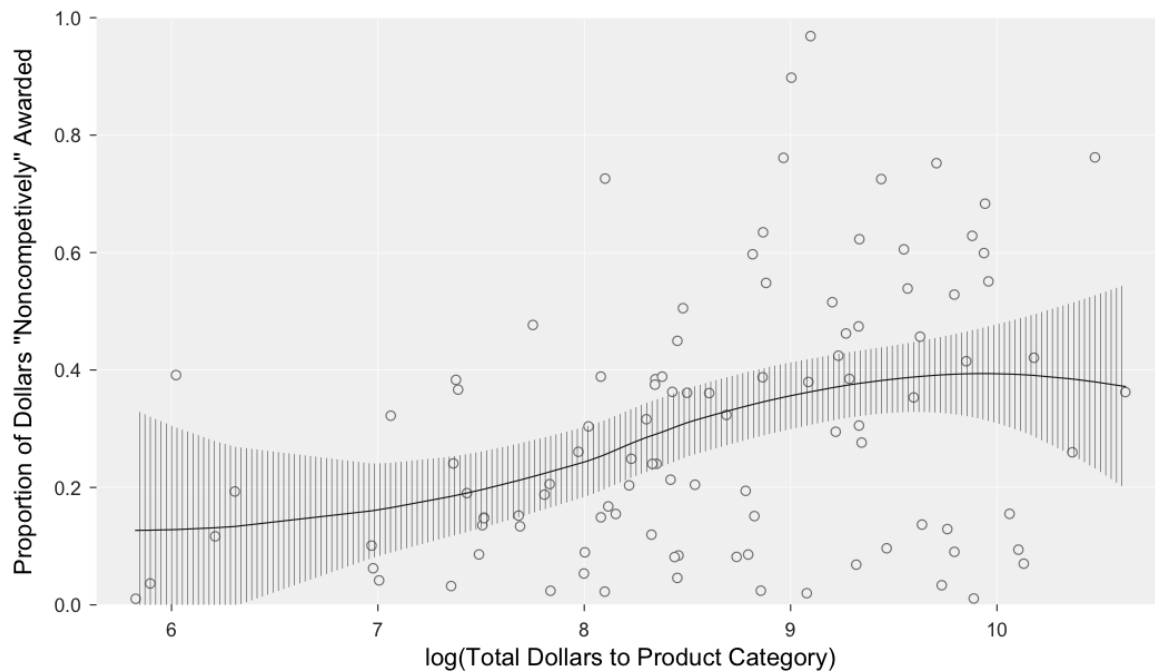
### *Model 2: BSTS with Traditional Synthetic Control as Potential Covariate*

Variable	9/11 Attacks				Troop Surge				Bin Laden Death			
	Mean	10%	50%	90%	Mean	10%	50%	90%	Mean	10%	50%	90%
Market <sub>t</sub>	0.04	0	0.01	0.07	0.39	0.02	0.15	1	0.47	0.05	0.37	1
PCA <sub>1</sub>	0.25	0.01	0.05	1	0.33	0.03	0.17	0.99	0.06	0.01	0.03	0.11
PCA <sub>2</sub>	0.36	0.01	0.11	1	0.23	0.02	0.12	0.68	0.32	0.02	0.15	0.95
PCA <sub>3</sub>	0.20	0.01	0.04	0.93	0.25	0.01	0.09	1	0.07	0.01	0.04	0.13
PCA <sub>4</sub>	0.17	0.01	0.06	0.55	0.14	0.01	0.07	0.33	0.11	0.02	0.05	0.25
PCA <sub>5</sub>	0.10	0.01	0.03	0.26	0.19	0.01	0.06	0.63	0.07	0.01	0.03	0.16
PCA <sub>6</sub>	0.07	0.01	0.02	0.14	0.10	0.02	0.05	0.22	0.07	0.01	0.03	0.16
PCA <sub>7</sub>	0.17	0.01	0.04	0.67	0.18	0.01	0.04	0.71	0.05	0	0.02	0.10
PCA <sub>8</sub>	0.08	0.01	0.02	0.17	0.13	0.01	0.06	0.33	0.06	0.01	0.03	0.12
PCA <sub>9</sub>	0.13	0.01	0.04	0.37	0.20	0.01	0.06	0.71	0.04	0.01	0.02	0.06
PCA <sub>10</sub>	0.29	0.01	0.09	1	0.15	0.01	0.06	0.40	0.03	0	0.01	0.05
Synth <sub>it</sub>	0.35	0.01	0.08	1	0.17	0.01	0.05	0.56	0.68	0.06	0.93	1

*Notes:* This table plots the frequency of various covariates being included in the final BSTS models. The upper panel presents results from the baseline model; the lower panel presents results from the BSTS model with traditional synthetic controls included. Cells denote average (or quantile) probabilities that individual covariates were selected by the spike-and-slab variable selection procedure (e.g., George and McCulloch, 1993; Ishwaran and Rao, 2005; Mitchell and Beauchamp, 1988). Quantiles are marked by the 10%, 50%, and 90% columns. The table demonstrates firm-level estimating equations varied markedly across event studies. For example, roughly 5 percent of firm-level models selected Market<sub>t</sub> as a covariate Model 1 in the 9/11 study. By contrast, about 40 percent of models selected Market<sub>t</sub> in the troop surge studies. The informativeness of the Synth<sub>it</sub> variable varies across event studies.

## A.5 Market Competition Across Defense Products

**Figure A.16:** *Higher-Earning Defense Products Less Likely to be “Competitively” Awarded (FY2000 to September 10, 2001)*



*Notes:* Each dot represents an individual, two-digit product category. The horizontal axis displays a product's total obligations between FY2000 and September 10, 2001, where total dollars are presented in log-10 units. The vertical axis indicates the share of dollars awarded “noncompetitively” within a given product category. The solid line is a local-regression line of best fit (with 95% confidence intervals surrounding the estimate). The graphic demonstrates there is an in-sample association between total obligations to a product category and the degree of market competition. Higher-earning product categories are less competitive, on average.

## A.6 Firm-Level Results Across Competing Estimators

### Differing Estimates via Time-Series Regression Discontinuity Design

Rather than estimate a structural model or synthetic controls, a possible estimation strategy in settings such as ours is via a regression discontinuity design (RDD). Lee and Lemieux (2010) provide a detailed overview of the method. RDD frameworks have gained enormous traction in recent years due to their interpretability and flexibility over a wide range of settings. In our applied context, as in Davis (2008), one can estimate a model in which time serves as the “forcing variable.” Estimates from the time-series RDD may lead to differing estimates from the BSTS or synthetic control paradigm, however. The argument here is that in the context of financial event studies—as financial returns are widely known to exhibit autocorrelation, and cross-sectional returns are known to be correlated within time periods—the possibility of period-specific common shocks could likely bias causal estimates.

For each exposed ticker in each event, utilizing the 20 trading days before and after the event, we estimate the following  $k$ -th order polynomial regressions:

$$Y_i = \mathbf{f}(T_i) + D_i \cdot \phi_i + \eta_i = \beta_0 + \left( \sum_k \beta_k t_i^k \right) + \mathbf{1}\{t > T_0\} \cdot \phi + \eta_i, \quad (5)$$

and robust confidence intervals are calculated as in Calonico, Cattaneo and Titiunik (2014). Equation 5 resembles that provided in Equation 1.2, but it explicitly assumes that the counterfactual series is well-approximated by the local-polynomial trend around the discontinuity (i.e., when  $t > T_0$ ). Beyond concerns about the correct polynomial order being specified (Gelman and Imbens, 2014) or the optimal bandwidth around the discontinuity, the design assumes the local trend appropriately maps to the counterfactual series in the post-event window. This assumption is not required in the BSTS or synthetic control paradigms utilized in the main body of this study.

To illustrate how the RDD in this context may lead to differing results from the BSTS

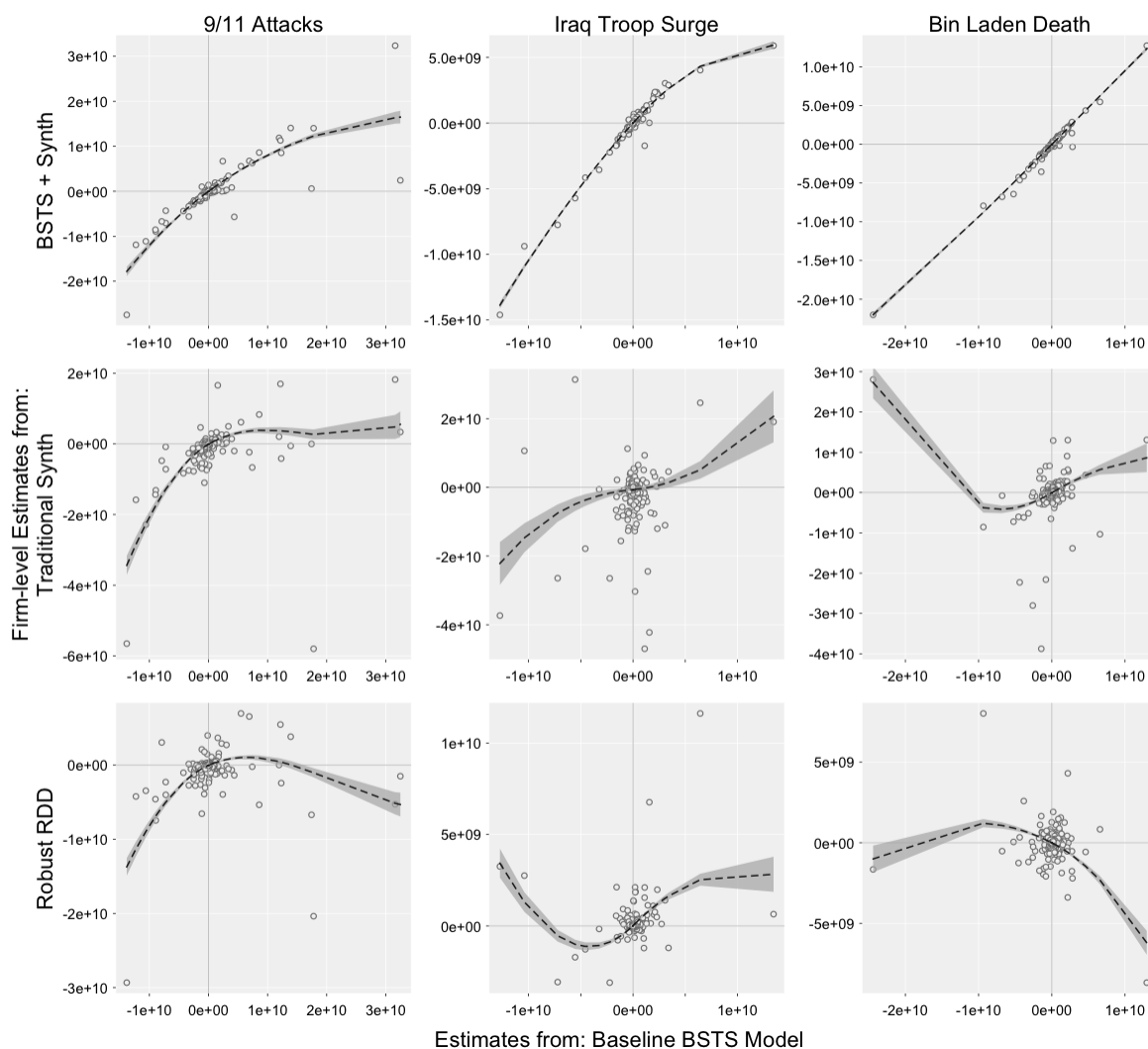
or synthetic control models, consider the following data generation processes:

$$Y_{it}^1 = \beta_0 + \left( \sum_t \beta_k t^k \right) + \gamma_t + \mathbf{1}\{t > T_0\} \cdot \phi + \varepsilon_{it}^1, \quad (6)$$

$$Y_{it}^0 = \beta_0 + \left( \sum_t \beta_k t^k \right) + \gamma_t + \varepsilon_{it}^0, \quad (7)$$

where  $\mathbb{E}(\gamma_t | t > T_0) - \mathbb{E}(\gamma_t | t \leq T_0) = \lambda$ . If one were to use Equation 5 to estimate  $\phi$ , but  $\phi$  is actually generated as in Equation 7, the expected bias of the RDD estimate is equal to  $\text{Bias}_{\text{rdd}} = \lambda + \phi - \mathbb{E}(Y_{it}^1 - Y_{it}^0 | t > T_0) = \lambda$ . This highlights the fact that period-specific common shocks may bias the RDD estimate in settings such as ours. In the time-specific common shocks are sufficiently large relative to  $\phi$ , it conceivable for the RDD estimate to have the “wrong sign” altogether. Figure A.17 shows how firm-level effects vary across event studies and estimators. Overall, estimates from the RDD ( $k=3$ ) are measurably different from those produced by the BSTS and traditional synthetic control estimators. Hence, due to a lack of correspondence with other estimators and on a priori grounds, we prefer the BSTS estimates over those from the time-series RDD.

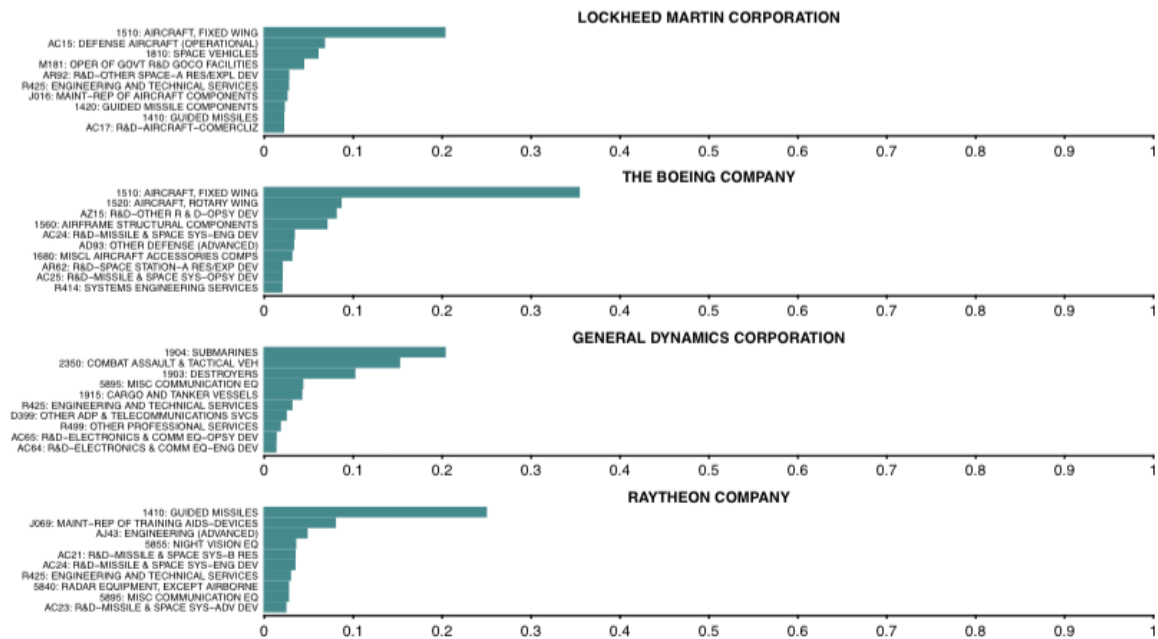
**Figure A.17: Relations Between Firm-Level Causal Estimates Across Competing Estimators and Events**



*Notes:* This graphic compares firm-level causal estimates across competing estimators and events. Dots mark causal estimates for individual firms, and the curved lines mark local-regression lines with 95% confidence intervals. BSTS and traditional synthetic control models are fit utilizing the 500 trading days before the event, and causal estimates are taken as in Equation 1.3.2 with a post-event window length of 20 trading days. RDD estimates are estimated using the 20 trading days before and after the event. There is a strong, positive correlation between firm-level estimates in the BSTS models; there is a weak but positive relationship between estimates derived from the BSTS models and the traditional synthetic control estimator; estimates derived from the Robust RDD as in Calónico, Cattaneo and Titiunik (2014) are weakly correlated the BSTS or synthetic control estimates.

## A.7 Supplemental Tables and Figures

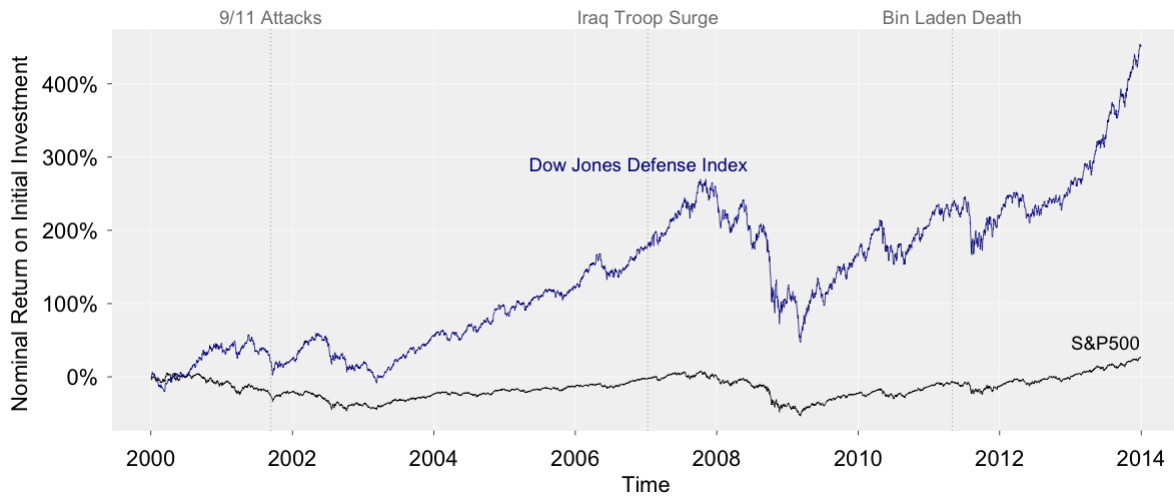
**Figure A.18:** Proportion of Products Sold to DoD by Firm (FY2009 to FY2011)



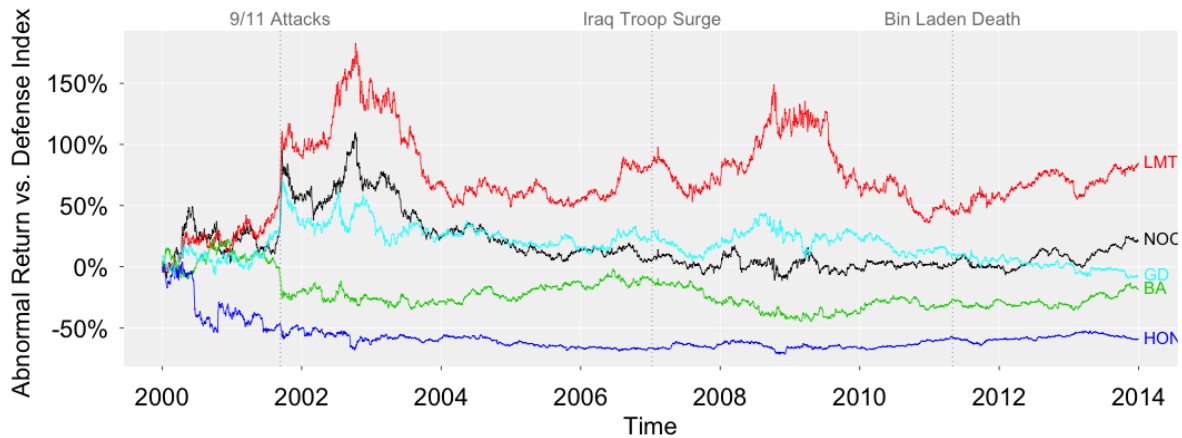
*Notes:* This figure shows which PSCs contribute most to each of the listed firms' total DoD revenues (for the highest grossing firms between FY2009 and FY2011, going from top to bottom). Within each subplot, the length of each bar is equal to the proportion of all DoD dollars earned by that PSC for that firm over the 2009 to 2011 interval. The figure highlights how top contract earners vary in terms of their product-level diversification.

**Figure A.19: Firm and Sector Returns vs. Benchmark Indices**

*(a) Dow Jones Defense Sector Returns vs. S&P500 Index*

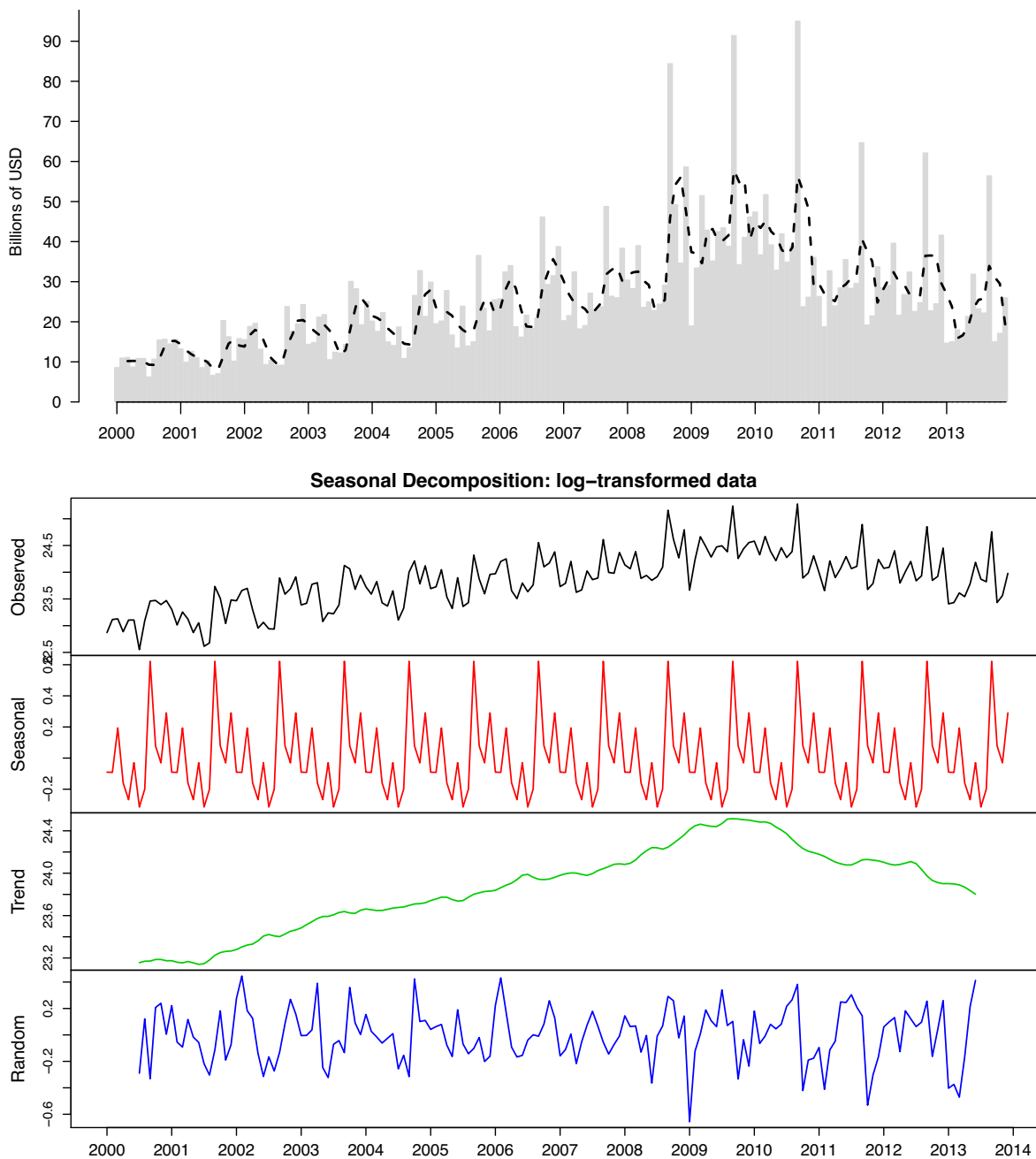


*(b) Firm Performance Relative to Defense Index*



*Notes:* The top panel shows the performance of the Dow Jones Defense Sector stock index against returns from the S&P500 over an equivalent interval.

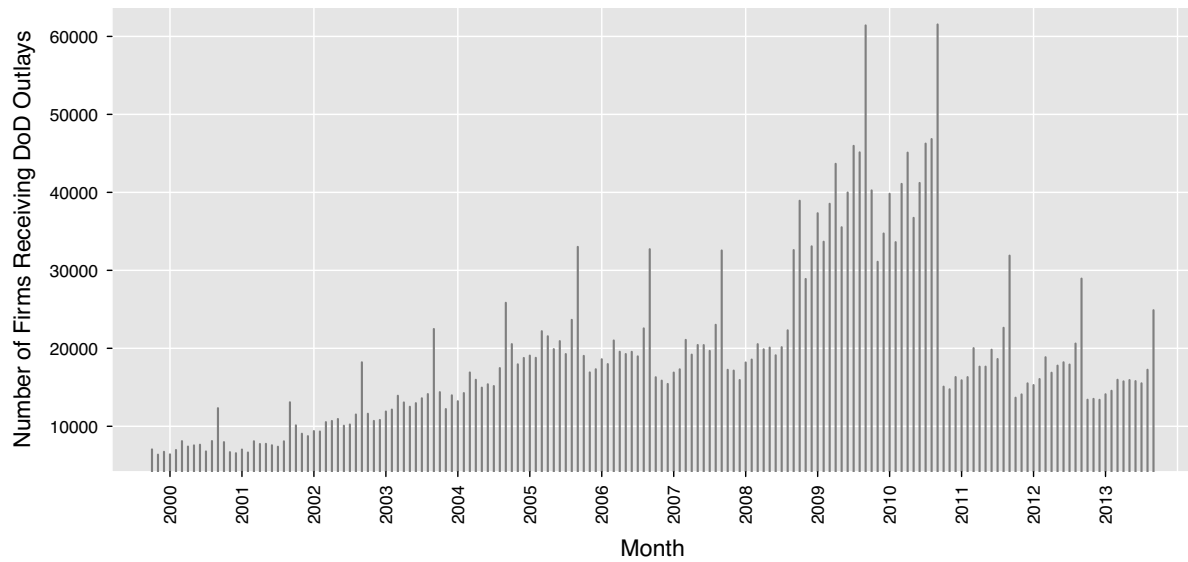
**Figure A.20: Monthly DoD Outlays (2000–2014)**



*Notes:* Data compiled from records obtained in the FPDS. Vertical bars denote total values of all monthly-level contract obligations at the United States Department of Defense. In the top panel, the dotted line reflects a three-month moving average. There is clear seasonality in the time series, as reflected in the lower panel. For example, at the end of each fiscal year (i.e., in September) there is a noticeable jump in allocations. The seasonal decomposition is given by  $\ln(Y_t) = T_t + S_t + e_t$ , where  $Y_t$  is the total monthly outlays in dollars,  $T_t$  is a trend modeled by a period-specific moving average, and  $e_t$  is the error. Results in the lower panel are in log-transformed units.



**Figure A.21:** Number of Firms Receiving DoD Outlays in each Month



*Notes:* This graphic shows growth in the unique number of firms receiving DoD outlays over a fourteen year span. The graphic demonstrates seasonality in the number of firms (e.g., a spike each year at the transition point between financial years), in addition to marked level changes in the 2008 to 2010 period. Since 2010, monthly totals resemble the 2005 to 2008 period, with a slight downward trend.

## B Appendix to Chapter 2

### B.1 Possible Data Limitations

To return to an issue raised in passing, readers may be concerned that with no 'top secret' cables in this particular sample, we are unable to obtain a true picture of diplomacy for this period. Our response is two fold: first, we assume that our understanding of the world is strictly increasing in the information we possess as researchers. That is, to the extent that we have the cables we do, we can learn more than we knew before: thus the study is worth undertaking. Second, from the perspective of the particular questions we ask below, it is unclear that the absence of more highly classified cables is a problem in terms of *bias*, specifically. Put otherwise, we believe our conclusions in what follows represent a 'lower bound': *if* certain matters are restricted in our sample (i.e. are classified as 'confidential' or 'secret') then we contend that these same types of issues would appear in 'top secret' cables, albeit to a larger extent. That is, we assume that secrecy is monotonic insofar as it is not the case that, for example, unclassified and 'top secret' cables are generated in fundamentally the same way, but that the intervening levels ('confidential' and 'secret') are completely different.

## B.2 How Cables are Written and Classified

U.S. State Department cables are official communications between Foreign Service officials, embassies, consulates, and international organizations (such as the United Nations Headquarters in New York City, The Hague) around the world. These communications serve at least two purposes. First, they exist to share information regarding the daily proceedings of an embassy or nation with other institutions around the world, on topics broadly related to the overall interests of the United States. Second, they exist to create official records (i.e., a database) of political information relevant to particular foreign outposts over time. An early articulation of these twinned objectives can be found in Department of State (1974, [http://aad.archives.gov/aad/content/aad\\_docs/rg59\\_state\\_dept\\_tags\\_74.pdf](http://aad.archives.gov/aad/content/aad_docs/rg59_state_dept_tags_74.pdf)).

When written, diplomatic cables take a standardized form. For example, the State Department provides glossaries of suggested language for officials to use in discussions of specific political issues in official communications. The “Termdex” chapter of 5 FAH-3 TAGS Terms Handbook serves this role, which “is an alphabetic list of words and phrases frequently found in Departmental communications”(1), requests officials use the terms “Border Incident” instead of “Border Violation” when discussing territorial disputes in official communications (5 FAH-3 H-810, pg. 9), or to use “Liberation Front” instead of “Liberation Movement” (5 FAH-3 H-810, pg. 40). So too, the Termdex recommends that particular phrases, if used, be used in tandem with specific subject TAGS—e.g., cables that discuss “Collective Bargaining” be tagged with the political tag “ELAB: Labor Sector Affairs”, and discussions of “Collective Security” be tagged with “MARR: Military and Defense Arrangements” (13).

After a cable is written and political subject TAGS are marked in metadata, cables are assigned an overall restriction status. Generally speaking, for any single cable in our sample it is impossible to the full set of individuals who may be involved in a single communication’s restriction status, just as it is impossible to know if esoteric

operational practices exist across U.S. embassies in the sample. That said, there is sufficient reason to believe a cable's restriction status is determined after a cable's text is complete, and generally this determination is made by a person of authority at the cable's location of origin (e.g., an Ambassador, or a Deputy Assistant Secretary, as articulated in "Original Classification Authority", 5 FAH-3 H-714.1, pg. 2). The Foreign Affairs Manual recommends a communication's "overall classification level is determined by the highest classification level of any of the portions" of its text. In other words, after a communication is written, an embassy's classification authority reviews all portions of a communication to check for the sensitivity of all portions of a message, where a "portion is ordinarily defined as a paragraph but also includes subject lines, titles, subheadings, tables, maps, photographs, graphs, and any other inserts within text" ("Classification Level", 5 FAH-3 H-713, pg. 2). This acknowledges that while not all portions of a cable may be equally sensitive, a cable's overall restriction status is set once all details of its contents have been reviewed. It is each cable's overall restriction status that we use for analysis in this study.

### B.3 Additional Details on Lasso and Random Forest

The RF and lasso procedures require brief explanation as they are not widely used in political research, although inevitably many technical details will be left for readers consult in the works cited. Both are widely used in “small  $n$ , large  $p$ ” settings: cases in which there may be there may be a greater number of possible parameters than observations in the sample.

The RF algorithm is a decision tree and resampling-based classification procedure which relies on repeatedly dividing the observed sample of data into random bootstrapped training datasets and fitting decision trees to each random training set, then aggregating the classification results over all independent training sets. In the ‘statistical learning’ literature, this procedure is commonly referred to as bootstrapped aggregation (i.e., “bagging”), and can be widely applied to improve the classification precision of various models, regression included. A RF algorithm procedure deviates from bagging alone by also randomly sampling the parameter space included in each iteration of this bagging procedure (e.g., Ho, 1998). One result of procedures like RF is it allows researchers to think about the relative *variable importance* of predictors in a classification setting. Due to the fact that at each bagged iteration of the procedure there are random subsets of the feature space included in the decision-trees, not all predicting variables (i.e., “words” in our context) are likely to be included as predictors at each stage of the algorithm. Overall, a predictor’s variable importance can be thought of as a result of this process: an estimate of the marginal reduction in classification error that results from a single word’s inclusion to the classification procedure overall, given the random inclusion of other predictor variables from the sample.

The lasso is a form of penalized regression, similar to ridge regression, whereby regression coefficients are weighted by “shrinkage factors” such that regression coefficients are weighted towards zero (Tibshirani, 1994; Hastie, Tibshirani and Friedman, 2009). The lasso is commonly used for feature selection in high-dimensional learning problems to

decrease the variance of a particular classifier. In our context, the procedure is similar to an ordinary least squares regression procedure in which the best-model is determined by that which minimizes the in-sample sum of squared residuals, except regression coefficients are penalized according to prior rules (i.e., the shrinkage factor and tuning factor) on the minimum coefficient size a variable is allowed to have to be included in the final classification model. The estimates presented in Figure 2.5 are obtained from taking the average lasso coefficient for each word over a representative range of shrinkage factors, where each model is estimated at the embassy level using its exactly matched data subset.

## B.4 Cable Tags

**Figure B.22:** State Department "Subject TAGS" in sample and meanings

	Meaning		Meaning
1) AADP	Automated Data Processing	51) ELTN	Land Transportation
2) ABLD	Buildings and Grounds	52) EMIN	Minerals and Metals
3) ABUD	Budget Services and Financial Systems	53) ENRG	Energy and Power
4) ACOA	Communication Operations and Administration	54) EPET	Petroleum and Natural Gas
5) ACKM	COMSEC Key Management	55) ETRD	Foreign Trade
6) ADCO	Diplomatic Courier Operations	56) ETTC	Trade and Technology Controls
7) ADPM	Diplomatic Pouch and Mail	57) EWWT	Waterborne Transportation
8) AEMR	Emergency Planning and Evacuation	58) MARR	Military and Defense Arrangements
9) AFIN	Financial Management	59) MASS	Military Assistance and Sales
10) AFSI	Foreign Service Institute	60) MCAP	Military Capabilities
11) AFSN	Foreign Service National Personnel	61) MNUC	Military Nuclear Applications
12) AGAO	General Accounting Office	62) MOPS	Military Operations
13) AINF	Information Management Services	63) ODIP	U.S. Diplomatic Representation
14) AINR	INR Program Administration	64) OEXC	Educational and Cultural Exchange Operations
15) AINT	Internet Administration	65) OFDP	Foreign Diplomats and Foreign Missions
16) ALOW	Allowances	66) OIIP	International Information Programs
17) AMED	Medical Services	67) OPDC	Diplomatic Correspondence
18) AMGT	Management Operations	68) OPRC	Public Relations and Correspondence
19) AMTC	Telecommunications Equipment Maintenance	69) OREP	U.S. Congressional Travel
20) ANET	Communications, Circuits, and Networks	70) OSCI	Science Grants
21) AODE	Employees Abroad	71) OTRA	Travel
22) AOMS	Office Management Specialist Issues	72) OVIP	Visits and Travel of Prominent Individuals and Leaders
23) AORC	International Organizations and Conferences	73) PARM	Arms Controls and Disarmament
24) APCS	Personal Computers	74) PBTS	National Boundaries, Territories, and Sovereignty
25) APER	Personnel	75) PGOV	Internal Governmental Affairs
26) ASCH	U.S. Sponsored Schools	76) PHSA	High Seas Affairs
27) ASEC	Security	77) PHUM	Human Rights
28) ASIG	Inspector General Activities	78) PINR	Intelligence
29) BBSR	Business Services Reporting	79) PINS	National Security
30) BEXP	Trade Expansion and Promotion	80) PNAT	National Independence
31) BMGT	FCS Management Operations	81) PREF	Refugees
32) BTIO	Trade and Investment Opportunities	82) PREL	External Political Relations
33) CASC	Assistance to Citizens	83) PROP	Propaganda and Psychological Operations
34) CFED	Federal Agency Services	84) PTER	Terrorists and Terrorism
35) CJAN	Judicial Assistance and Notarial Services	85) SCUL	Cultural Affairs
36) CLOK	Visa Lookout	86) SENV	Environmental Affairs
37) CMGT	Consular Administration and Management	87) SMIG	Migration
38) CPAS	Passport and Citizenship	88) SNAR	Narcotics
39) CVIS	Visas	89) SOCI	Social Conditions
40) EAGR	Agriculture and Forestry	90) TBIO	Biological and Medical Science
41) EAID	Foreign Economic Assistance	91) TINT	Internet Technology
42) EAIR	Civil Aviation	92) TNGD	Engineering Research and Development
43) ECON	Economic Conditions	93) TPHY	Physical Sciences
44) ECPS	Communications and Postal Systems	94) TRGY	Energy Technology
45) EFIN	Financial and Monetary Affairs	95) TSPA	Space Activities
46) EFIS	Commercial Fishing and Fish Processing	96) TSPL	Science and Technology Policy
47) EIND	Industry and Manufacturing		
48) EINT	Economic and Commercial Internet		
49) EINV	Foreign Investments		
50) ELAB	Labor Sector Affairs		

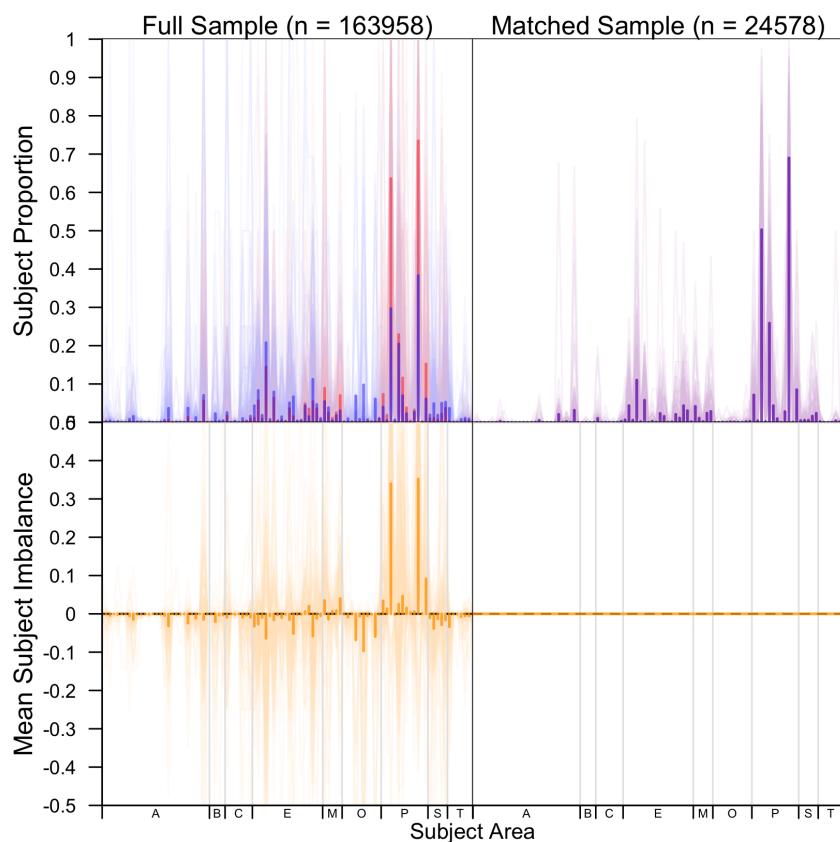
Notes: Meanings taken from "Subject TAGS" in the U.S. Department of State Foreign Affairs Manual Volume 5, Handbook 3.





## B.6 Reduction in Imbalance from Matching

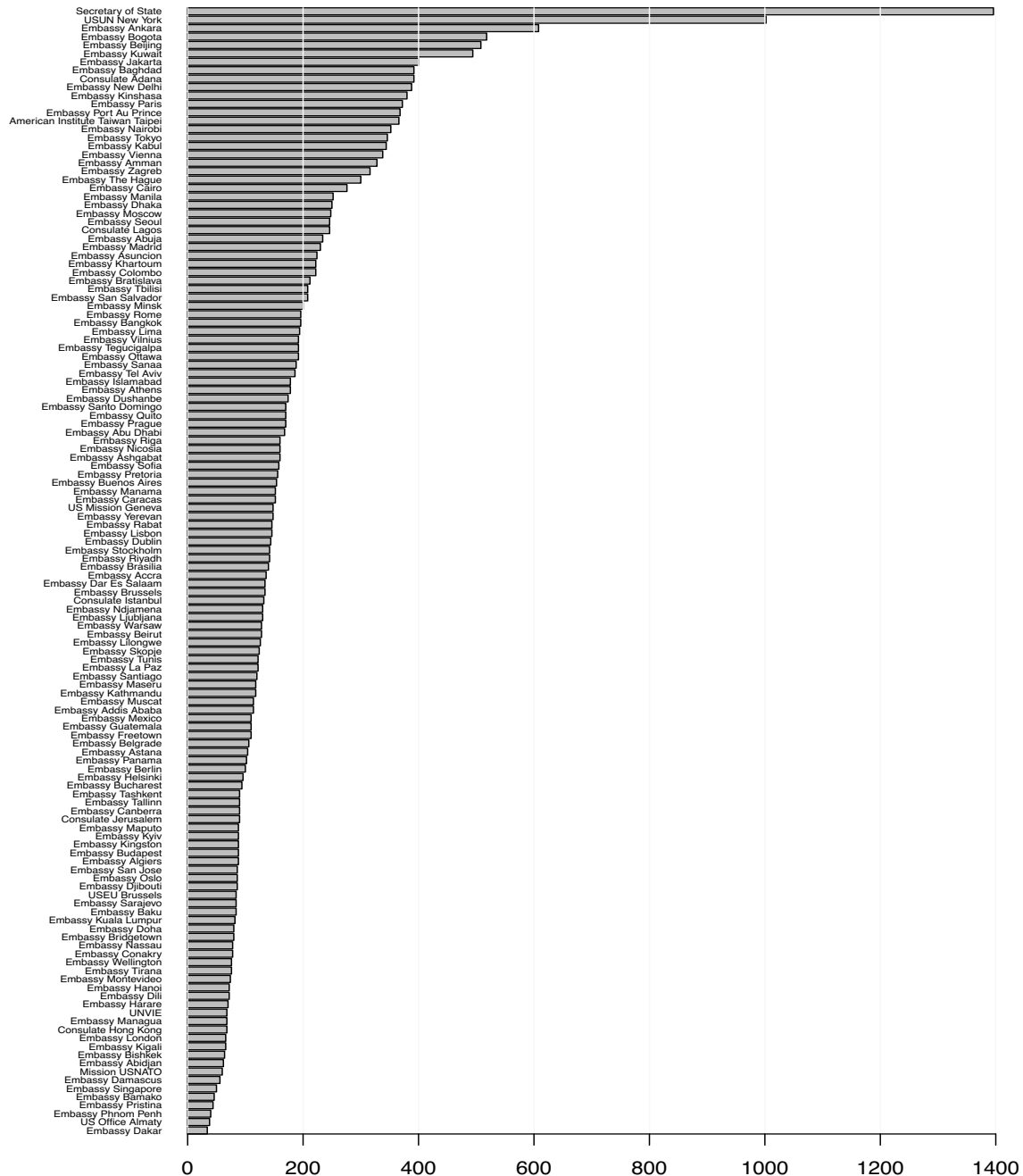
**Figure B.24:** *Reduction in Subject Imbalance From Exact Matching*



*Notes:* In the upper-left plot, the lines in red (restricted) and blue (unrestricted) correspond to embassy-level averages of tag frequencies of respective classification levels. Thicker lines in the foreground denote sample averages. Axis labels of individual subject tags are omitted for visual clarity and instead labeled with a broader subject tag. The lower-left plot presents this in the form of tag-level imbalance: background lines indicate embassy-level imbalances (subject tag differences in means between restricted and unrestricted cables by embassy) while bars in the foreground are the sample differences. This demonstrates there is subject imbalance between unrestricted and restricted cables both on aggregate and at embassy levels. In the study sample, categories like *A – Administrative Affairs*, *B – Business Services*, *E – Economic Affairs*, *O – Outreach* tend to be less classified on average, whereas *M – Military and Defense Affairs* and *P – Political Affairs* tend to be more private. Exact matching perfectly reduces within-embassy and across-embassy tag imbalance.

## B.7 Counts by Embassy in Matched Sample

Figure B.25: Counts of cables by embassy in the matched sample.



Notes: Horizontal bars denote total cables included (per embassy) in the matched study sample.

## B.8 Exact Matching Algorithm

**Figure B.26:** *Outline of Exact Matching Algorithm*

1. Let  $N_j$  be the set of cables from embassy  $j$  that occur during or after the year 2005 in the sample, where  $|N_j|$  is the number of cables originating from location  $j$ .
2. For each of the  $|N_j|$  documents in the sample, record the subject tags present on each diplomatic cable.
3. For all restricted cables in  $N_j$ , find all unrestricted cables in  $N_j$  that exactly match on subject tags and year of creation.
4. From the subset of restricted cables in  $N_j$  with at least one unrestricted exact match,
  - (a) Randomly draw a restricted cable and find the unrestricted, exact-matched cable that is written most closely in time (i.e., the cable that minimizes the absolute value between the difference in release days). Each cable may be matched with or without replacement.
  - (b) Continue this process until there are no-more restricted and unrestricted cables to pair together.
5. Record the list of exactly-matched pairs of cables, if applicable.

## B.9 Probabilistic Topic Model

The field of quantitative text analysis has grown substantially in recent years. In this literature, applied researches extensively use Latent Dirichlet Allocation (Blei, Ng and Jordan, 2003) as a generative model to extract “themes” or “topics” from a collection of documents. See, for example, Blei (2012) for an overview, and Quinn et al. (2010) for recent political science applications of topic models. The model assumes documents are composed of latent topics that are chosen with probabilities following a Dirichlet distribution, and multinomial choice probabilities for word choice conditional on a topic. More precisely, the framework from Blei, Ng and Jordan (2003) has the number of words  $N$  in a document be  $\text{Poisson}(\xi)$ , the latent topic probabilities  $\theta$  be  $\text{Dirichlet}(\alpha)$ , the topics  $z_n$  be  $\text{Multinomial}(\theta)$ , and the words  $w_n$  be  $\text{Multinomial}(\beta)$ , conditional on  $z_n$ . Then, with  $M$  documents, they have that  $p(C|\alpha, \beta) = \prod_{d=1}^M \int p(\theta_d|\alpha) \left( \prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta) \right) d\theta_d$ . Computational difficulties arise in this setting, but there are ways to deal with them (e.g., Hoffman et al., 2013).

When a researcher estimates an LDA model, the topics returned are characterized by the multinomial probabilities for all words within each topic, as well as the posterior distribution of topics conditional on a certain word. In practice, the researcher selects the number of topics a priori, although recent efforts have been made to assess how an approximate number of topics may be present in a sample of data (see, e.g., Hoffman et al., 2013). We used these posterior estimates to generate the topical ordering in Figure 2.5: for each topic, we took the cosine similarity (e.g., Manning, Raghavan and Schütze, 2008) between all pairs of topics, where similarity between topic vectors is determined by the posterior weights placed on each word in each topic.

## C Appendix to Chapter 3

### C.1 Derivations of Growth in Similarity Measures

Suppose we have pre- and post- transcripts (noted, respectively, as vectors  $p_0$  and  $p_1$ ) and press releases ( $q_0$  and  $q_1$ ). Our aim here is to derive growth measures for *cosine similarity* and *generalized cosine similarity* that allow us to decompose document-level and word-level contributions to observed growth.

#### Cosine Similarity

We begin with

$$s_0 = \frac{p_0' q_0}{\|p_0\| \|q_0\|},$$

and

$$s_1 = \frac{p_1' q_1}{\|p_1\| \|q_1\|}.$$

The growth of the similarity measure is defined as

$$\frac{s_1 - s_0}{s_0},$$

which in continuous time would be

$$\frac{\frac{\partial s}{\partial t}}{s_0},$$

or

$$\frac{\partial \ln(s)}{\partial t}.$$

The quantity  $\ln(s_1) - \ln(s_0)$  would be a discrete time way to measure the “growth” of the similarity measure over time. Note that in our applied setting we are interested in the effect of a policy change. In this case,  $\ln(s_2) - \ln(s_0)$  would also be a measure of growth.

The derivation below will show how to decompose the growth rates by word and by document, in order to answer the following counterfactual: Suppose that we imposed  $p_{1j} = p_{0j}$  for some stem  $j$ . How much would change the growth in similarity?

We start with

$$\begin{aligned}
\frac{\partial \ln(s)}{\partial t} &= \frac{\partial \ln\left(\frac{p'q}{\|p\|\|q\|}\right)}{\partial t} = \frac{\partial \ln(p'q)}{\partial t} - \frac{\partial \ln(\|p\|\|q\|)}{\partial t} \\
&= \frac{\partial \ln(\sum p_j q_j)}{\partial t} - \frac{1}{2} \left( \frac{\partial \ln(\sum p_j^2)}{\partial t} + \frac{\partial \ln(\sum q_j^2)}{\partial t} \right) \\
&= \sum \left( \frac{\partial p_j}{\partial t} \frac{q_{0j}}{\sum p_{0j} q_{0j}} + \frac{\partial q_j}{\partial t} \frac{p_{0j}}{\sum p_{0j} q_{0j}} - \frac{\partial p_j}{\partial t} \frac{p_{0j}}{\sum p_{0j}^2} - \frac{\partial q_j}{\partial t} \frac{q_{0j}}{\sum q_{0j}^2} \right) \\
&= \sum \left( \frac{\frac{\partial p_j}{\partial t} p_{0j} q_{0j}}{p_{0j} \sum p_{0j} q_{0j}} + \frac{\frac{\partial q_j}{\partial t} p_{0j} q_{0j}}{q_{0j} \sum p_{0j} q_{0j}} - \frac{\frac{\partial p_j}{\partial t} p_{0j}^2}{p_{0j} \sum p_{0j}^2} - \frac{\frac{\partial q_j}{\partial t} q_{0j}^2}{q_{0j} \sum q_{0j}^2} \right) \\
&\approx \sum \left( \frac{\partial \ln(p_j)}{\partial t} \frac{p_{0j} q_{0j}}{\sum p_{0j} q_{0j}} + \frac{\partial \ln(q_j)}{\partial t} \frac{p_{0j} q_{0j}}{\sum p_{0j} q_{0j}} - \frac{\partial \ln(p_j)}{\partial t} \frac{p_{0j}^2}{\sum p_{0j}^2} - \frac{\partial \ln(q_j)}{\partial t} \frac{q_{0j}^2}{\sum q_{0j}^2} \right) \\
&= \sum \left( \frac{\partial \ln(p_j)}{\partial t} \left[ \frac{p_{0j} q_{0j}}{\sum p_{0j} q_{0j}} - \frac{p_{0j}^2}{\sum p_{0j}^2} \right] \right) \tag{8} \\
&\quad + \sum \left( \frac{\partial \ln(q_j)}{\partial t} \left[ \frac{p_{0j} q_{0j}}{\sum p_{0j} q_{0j}} - \frac{q_{0j}^2}{\sum q_{0j}^2} \right] \right) \tag{9}
\end{aligned}$$

where the sums are over all of the stems  $j$ . If we want to look at  $p_j$ 's effect on the growth rate, we just calculate line 8 for stem  $j$ .

This also allows us to decompose sources and sinks. There are four options for line 8:

$$\begin{aligned}
\frac{\partial \ln(p_j)}{\partial t} > 0, \left[ \frac{p_{0j} q_{0j}}{\sum p_{0j} q_{0j}} - \frac{p_{0j}^2}{\sum p_{0j}^2} \right] > 0 &\implies \text{similarity decrease by increasing over-represented words,} \\
\frac{\partial \ln(p_j)}{\partial t} < 0, \left[ \frac{p_{0j} q_{0j}}{\sum p_{0j} q_{0j}} - \frac{p_{0j}^2}{\sum p_{0j}^2} \right] < 0 &\implies \text{similarity decrease by increasing under-represented words,} \\
\frac{\partial \ln(p_j)}{\partial t} > 0, \left[ \frac{p_{0j} q_{0j}}{\sum p_{0j} q_{0j}} - \frac{p_{0j}^2}{\sum p_{0j}^2} \right] < 0 &\implies \text{similarity increase by increasing under-represented words,} \\
\frac{\partial \ln(p_j)}{\partial t} < 0, \left[ \frac{p_{0j} q_{0j}}{\sum p_{0j} q_{0j}} - \frac{p_{0j}^2}{\sum p_{0j}^2} \right] > 0 &\implies \text{similarity increase by decreasing over-represented words.}
\end{aligned}$$

Similarly, we can decompose this quantity for  $q$ .

## Generalized Cosine Similarity

Now the similarity measure is

$$CS_{\Omega}(d_1, d_2) = \frac{\langle d_1, d_2 \rangle_{\Omega}}{\|d_1\|_{\Omega} \cdot \|d_2\|_{\Omega}} = \frac{p' \cdot \Omega \cdot q}{\sqrt{(p' \cdot \Omega \cdot p)} \times \sqrt{(q' \cdot \Omega \cdot q)}}$$

so

$$\begin{aligned} \frac{\partial \ln(CS_{\Omega})}{\partial t} &= \frac{\partial \ln\left(\frac{p' \cdot \Omega \cdot q}{\sqrt{(p' \cdot \Omega \cdot p)} \times \sqrt{(q' \cdot \Omega \cdot q)}}\right)}{\partial t} = \frac{\partial \ln(p' \cdot \Omega \cdot q)}{\partial t} - \frac{\partial \ln\left(\sqrt{(p' \cdot \Omega \cdot p)} \times \sqrt{(q' \cdot \Omega \cdot q)}\right)}{\partial t} \\ &= \frac{\partial \ln\left(\sum_j (p_j \cdot (\sum_i \omega_{ij} q_i))\right)}{\partial t} - \frac{1}{2} \left( \frac{\partial \ln\left(\sum_j (p_j \cdot (\sum_i \omega_{ij} p_i))\right)}{\partial t} + \frac{\partial \ln\left(\sum_j (q_j \cdot (\sum_i \omega_{ij} q_i))\right)}{\partial t} \right) \\ &= \sum \left( \frac{\partial p_j}{\partial t} \frac{\sum_i \omega_{ij} q_i}{\sum_j (p_j \cdot (\sum_i \omega_{ij} q_j))} + \frac{\partial q_j}{\partial t} \frac{\sum_i \omega_{ji} p_i}{\sum_j (p_j \cdot (\sum_i \omega_{ji} q_j))} \right. \\ &\quad \left. - \frac{\partial p_j}{\partial t} \frac{\sum_i \omega_{ij} p_i}{\sum_j (p_j \cdot (\sum_i \omega_{ij} p_j))} - \frac{\partial q_j}{\partial t} \frac{\sum_i \omega_{ij} q_i}{\sum_j (q_j \cdot (\sum_i \omega_{ij} q_j))} \right) \\ &= \sum \left( \frac{\frac{\partial p_j}{\partial t}}{p_j} \frac{p_j \sum_i \omega_{ij} q_i}{\sum_j (p_j \cdot (\sum_i \omega_{ij} q_j))} + \frac{\frac{\partial q_j}{\partial t}}{q_j} \frac{q_j \sum_i \omega_{ji} p_i}{\sum_j (p_j \cdot (\sum_i \omega_{ji} q_j))} \right. \\ &\quad \left. - \frac{\frac{\partial p_j}{\partial t}}{p_j} \frac{p_j \sum_i \omega_{ij} p_i}{\sum_j (p_j \cdot (\sum_i \omega_{ij} p_j))} - \frac{\frac{\partial q_j}{\partial t}}{q_j} \frac{q_j \sum_i \omega_{ij} q_i}{\sum_j (q_j \cdot (\sum_i \omega_{ij} q_j))} \right) \\ &= \sum \left( \frac{\partial \ln(p_j)}{\partial t} \frac{p_j \sum_i \omega_{ij} q_i}{\sum_j (p_j \cdot (\sum_i \omega_{ij} q_j))} + \frac{\partial \ln(q_j)}{\partial t} \frac{q_j \sum_i \omega_{ji} p_i}{\sum_j \left( p_j \cdot \left( \sum_i \omega_{ji} q_j \right) \right)} \right. \\ &\quad \left. - \frac{\partial \ln(p_j)}{\partial t} \frac{p_j \sum_i \omega_{ij} p_i}{\sum_j (p_j \cdot (\sum_i \omega_{ij} p_j))} - \frac{\partial \ln(q_j)}{\partial t} \frac{q_j \sum_i \omega_{ij} q_i}{\sum_j (q_j \cdot (\sum_i \omega_{ij} q_j))} \right) \\ &= \sum \left( \frac{\partial \ln(p_j)}{\partial t} \left[ \frac{p_j \sum_i \omega_{ij} q_i}{\sum_j (p_j \cdot (\sum_i \omega_{ij} q_j))} - \frac{p_j \sum_i \omega_{ij} p_i}{(p_j \cdot (\sum_i \omega_{ij} p_j))} \right] \right) \quad (10) \\ &\quad + \sum \left( \frac{\partial \ln(q_j)}{\partial t} \left[ \frac{q_j \sum_i \omega_{ji} p_i}{\sum_j (p_j \cdot (\sum_i \omega_{ji} q_j))} - \frac{q_j \sum_i \omega_{ij} q_i}{(q_j \cdot (\sum_i \omega_{ij} q_j))} \right] \right) \quad (11) \end{aligned}$$

with a similar decomposition as before. What matters now, however, is the weighted average contribution.

## C.2 Topic modeling with a dictionary

In this section we consider generating economics clusters (or “topics”) in the dictionary, as another way of identifying what types of language was most responsible for the

increase in similarity. Running our analysis at the dictionary level allows our results to be replicated in other settings, since the particular set and order of topics will not change. It also has the benefit of not building a measure of word relations that is sensitive to the specific equilibrium of a study sample.

To supplement the analysis found in the main body of our paper, we implement a variant of Latent Dirichlet Allocation (Blei, Ng and Jordan, 2003) on the ODE document-term matrix constructed from the dictionary's definitions. Numerous extensions to LDA have been used extensively in the text analysis literature to estimate "latent" topics from a set of documents given observed collocations between words in documents. The generic model has latent topics that are chosen with probabilities following a Dirichlet distribution, and multinomial choice probabilities for word choice conditional on a topic. The estimated model gives us, among other things, the multinomial probabilities for all words within each topic, as well as the posterior distribution of topics conditional on a certain word. More precisely, the setup from Blei, Ng and Jordan (2003) has the corpus  $\mathcal{C}$ , the number of words  $N$  in a document be Poisson( $\xi$ ), the latent topic probabilities  $\theta$  be Dirichlet( $\alpha$ ), the topics  $z_n$  be Multinomial( $\theta$ ), and the words  $w_n$  be Multinomial( $\beta$ ), conditional on  $z_n$ . Then, with  $M$  documents, they have the likelihood

$$p(\mathcal{C}|\alpha, \beta) = \prod_{d=1}^M \int p(\theta_d|\alpha) \left( \prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta) \right) d\theta_d$$

Difficulties arise in estimating parameters from this model, and we follow suggestions in line with Blei (2013) for computation.

We estimate a relational topic model (RTM) on the document-term matrix assembled from the Oxford Dictionary of Economics—a method introduced in Chang and Blei (2009, 2010). Unlike LDA, the RTM assumes there is a network structure to the documents in a corpus, and that "links" (or "edges") between documents are reflective of underlying similarity in document content. The observed document-network structure is used as a part of a "link prediction" problem (e.g., Liben-Nowell and Kleinberg, 2003). Documents



are first created in line with the LDA data-generating process. Pair-wise links are then modeled as logistic regressions conditional on the posterior topic distributions in each document. Chang and Blei (2010) provides additional detail on this procedure.

In our applied context, we estimate the RTM using 60 topics, with both of the model's prior concentration parameters set to 0.1. Each entry in the ODE—e.g., the definition for the word *inflation*—is treated as an individual document, and links between documents are determined by observed cross-references between dictionary entries. Figure C.27 shows the the 5 words with the highest posterior probabilities are presented in each of the 60 topics estimated in the dictionary. For each word stem present in the ODE, this procedure yields a vector of posterior topic probabilities given an occurrence of that word. We utilize this collection of estimated posterior topic probabilities an alternative approach to the construction of the weight matrix  $\Omega$ . To generate the weights using results from the RTM, we set each diagonal entry of  $\Omega$  to 1, we define each off-diagonal entry  $\Omega_{ij} = (\mathbf{p}_i \cdot \mathbf{p}_j)/10$ , where  $\mathbf{p}_i$  is the vector of (estimated) posterior topic probabilities given word  $i$ . The dot-product of these vectors has a natural interpretation given the generative model: it is an estimate of the probability that a random occurrence of  $i$  and random occurrence of  $j$  are drawn from the same latent topic. We down-weight this quantity by taking the square to adjust for the fact that even if two words are always drawn from the same topic, this does not imply that the words are perfect substitutes or synonyms.

The core results of our analysis barely change when the word-to-word weight matrices are constructed with the procedure described in this section. This is not ex-post surprising because there is a strong and positive association (correlation  $\approx 0.99$ ) between GCS values at the meeting level generated using our two methods of calculating semantic similarity. We leave improvements to the generation of this matrix—such as how to aggregate specialized dictionaries with existing lexical taxonomies, such as WordNet—as a task for future research.

### C.3 A bound on bias in similarity when trimming vectors

In this section, we provide a proof that dropping rare words will have a limited effect on our results, while increasing computation efficiency substantially. As a supplement to the results presented in Figure 3.2, this bound justifies our ex ante dimension reduction in the analysis (i.e., moving from over 34,000 word-stems present in the original public and private documents, to the study sample of 4,030 stems that intersect with the ODE).

*Proposition 1:* Suppose  $x = (x'_1, x'_2)'$  and  $y = (y'_1, y'_2)'$  are vectors in  $\mathbb{R}_+^n$ , where  $\|x\| \geq \|x_1\| \geq c\|x\|$  and  $\|y\| \geq \|y_1\| \geq c\|y\|$ . Then  $|CS(x, y) - CS(x_1, y_1)| \leq 1 - c^2$ .

*Proof:* First, note that  $\|x_2\| \leq \sqrt{1 - c^2}\|x\|$  and  $\|y_2\| \leq \sqrt{1 - c^2}\|y\|$ . It follows that  $|CS(x, y) - CS(x_1, y_1)| = \left| \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|} - \frac{\langle x_1, y_1 \rangle}{\|x_1\| \cdot \|y_1\|} \right| = \left| \frac{\langle x_1, y_1 \rangle + \langle x_2, y_2 \rangle}{\|x\| \cdot \|y\|} - \frac{\langle x_1, y_1 \rangle}{\|x_1\| \cdot \|y_1\|} \right| = \left| \frac{\|x_1\| \cdot \|y_1\| \cos(\theta_1) + \|x_2\| \cdot \|y_2\| \cos(\theta_2)}{\|x\| \cdot \|y\|} - \frac{\|x_1\| \cdot \|y_1\| \cos(\theta_1)}{\|x_1\| \cdot \|y_1\|} \right| = \left| \frac{(\|x_1\| \cdot \|y_1\| - \|x\| \cdot \|y\|) \cos(\theta_1) + \|x_2\| \cdot \|y_2\| \cos(\theta_2)}{\|x\| \cdot \|y\|} \right| \leq \max\left\{ \frac{(\|x\| \cdot \|y\| - \|x_1\| \cdot \|y_1\|)}{\|x\| \cdot \|y\|} \cos(\theta_1), \frac{\|x_2\| \cdot \|y_2\|}{\|x\| \cdot \|y\|} \cos(\theta_2) \right\} \leq \max\{(1 - c^2) \cos(\theta_1), (1 - c^2) \cos(\theta_2)\} \leq 1 - c^2$ .

*Proposition 2:* Suppose  $x = (x'_1, x'_2)'$  and  $y = (y'_1, y'_2)'$  are vectors in  $\mathbb{R}_+^n$ , where  $\|x_1\| = c\|x\|$  and  $\|y_1\| = c\|y\|$ . Then  $|CS(x, y) - CS(x_1, y_1)| = |(1 - c^2)(CS(x_2, y_2) - CS(x_1, y_1))|$ .

*Proof:* First, note that  $\|x_2\| = \sqrt{1 - c^2}\|x\|$  and  $\|y_2\| = \sqrt{1 - c^2}\|y\|$ .  $|CS(x, y) - CS(x_1, y_1)| = \left| \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|} - \frac{\langle x_1, y_1 \rangle}{\|x_1\| \cdot \|y_1\|} \right| = \left| \frac{\langle x_1, y_1 \rangle + \langle x_2, y_2 \rangle}{\|x\| \cdot \|y\|} - \frac{\langle x_1, y_1 \rangle}{\|x_1\| \cdot \|y_1\|} \right| = \left| \frac{\|x_1\| \cdot \|y_1\| \cos(\theta_1) + \|x_2\| \cdot \|y_2\| \cos(\theta_2)}{\|x\| \cdot \|y\|} - \frac{\|x_1\| \cdot \|y_1\| \cos(\theta_1)}{\|x_1\| \cdot \|y_1\|} \right| = |(c^2 - 1) \cos(\theta_1) + (1 - c^2) \cos(\theta_2)| = |(1 - c^2)(CS(x_2, y_2) - CS(x_1, y_1))|$ .

**Figure C.27: RTM topics in the ODE**

---

1. instrume, polici, target, announc, fiscal	31. mobil, unemploy, search, migrat, immigr
2. criterio, sex, afford, race, religion	32. auction, deliveri, hedg, spot, bidder
3. rent, tenant, land, occupi, landlord	33. voter, axiom, altern, vote, rank
4. budget, deficit, propens, inject, leakag	34. regress, squar, explanat, correl, heterosc
5. willing, partner, integr, franchis, merger	35. london, nyse, minut, york, 500
6. discount, matur, redeem, bill, gilt	36. durabl, car, furnitur, food, club
7. inequ, donor, lorenz, aid, core	37. levi, tax, taxpay, taxabl, vat
8. weight, herfinda, laspeyr, paasch, proxi	38. local, grant, defens, public, feder
9. asymmetr, inform, incomple, reveal, signal	39. statemen, thing, necessar, true, suffici
10. children, poverti, parent, famili, univers	40. imf, war, oil, wood, ibrd
11. inflatio, inflat, keynesia, spiral, pressur	41. capac, recov, spare, believ, threat
12. insur, scheme, life, pension, death	42. elast, slope, substitu, curv, downward
13. parti, arbitr, counter, settl, swap	43. deposit, bank, merchant, debit, notic
14. indiffer, util, solut, ordin, cardin	44. alloc, pareto, box, debreu, effici
15. qualiti, discrimi, resal, advertis, brand	45. null, likeliho, hypothes, test, asymptot
16. trust, investor, takeov, issu, isa	46. uruguay, round, multilat, talk, gatt
17. hour, worker, dismiss, redund, employe	47. float, convert, dollar, currenc, pariti
18. innov, patent, technic, knowledg, disembod	48. game, player, strategi, confess, play
19. entri, entrant, cournot, monopoli, rival	49. grow, growth, human, solow, capita
20. export, import, dump, devalu, intra	50. seri, stationa, autocorr, infin, root
21. pollut, emiss, global, atmosphe, pecuniar	51. survey, execut, board, review, 1997
22. deadweig, abat, margin, accru, implicit	52. input, output, intens, factor, isoqu
23. regul, safeti, antitrus, commiss, cartel	53. see, admit, park, caribbea, compris
24. gambl, odd, neutral, avers, prospect	54. franc, european, netherla, belgium, itali
25. arbitrag, portfoli, project, idiosync, return	55. resid, abroad, inward, subtract, visibl
26. analysi, behavior, studi, microeco, analyz	56. dividend, debentur, syndic, sharehol, unlimit
27. auditor, professi, record, audit, figur	57. eastern, soviet, sector, plan, mine
28. cycl, recess, quit, path, boom	58. sheet, deprec, obsolesc, wear, revalu
29. debtor, creditor, lender, insolv, unsecur	59. phillip, run, long, short, transito
30. rpi, gnp, deflat, volum, index	60. coinag, coin, intrins, token, refus

---

*Notes:* This figure shows, within each topic, the five word-stems with the highest posterior probability of assignment to that topic given an utterance of that word. Estimates are obtained from the Relational Topic Model algorithm (Chang and Blei, 2009, 2010) on the Oxford Dictionary of Economics, where the model was estimated with a total of 60 topics. For presentational purposes in this figure, word-stems were condensed to have a character length of no more than 8. The numeric identifier for each topic is arbitrary.

In the data, we calculate the  $c$  for each meeting from using dictionary-matching word stems instead of all stems. The average  $c$  is 0.98.

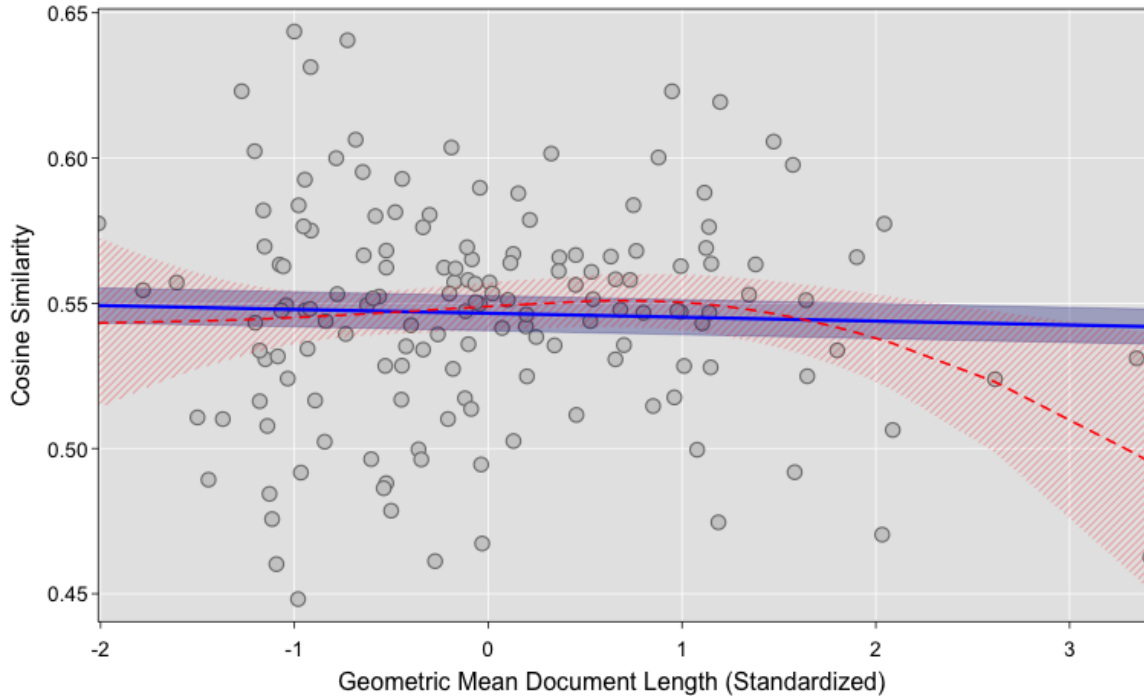
## C.4 Effects of Sample Size

In this section, we discuss the mechanical effects that the length of the documents may have on similarity (Gentzkow, Shapiro and Taddy, 2015). The intuition for the concern is as follows: for a given document length, as the number of dimensions increase, so too will the variance of the count for each dimension. This will mechanically lower estimated similarity.

In order to study how important this effect is in our setting, we run two tests. In Figure C.28, we study the observed correlation in the pre-period of document length and similarity. On the  $x$ -axis we plot the (standardized) geometric mean of the length of the public and private documents, and on the  $y$ -axis we plot the similarity. A one standard-deviation increase in average document length is predicted to decrease document similarity by 0.001. This suggests that, over the range of document lengths in the sample, more words does not particularly cause more similarity.

In Figure C.29, we undertake simulations under the following thought experiment. Suppose each document were a draw from an underlying probability distribution for each word, where each document type in each period has its own multinomial distribution. Each word's probability comes from its share of usage for its document type in its period. The left panel plots the mean change in similarity for each meeting, and the right panel plots the change in the underlying "treatment" effect. For cosine similarity, the standard deviation of estimated change is 0.002, and the mean is -0.003, so the mean simulated similarity understates the observed effect by -2.6%. For generalized cosine similarity, the mean simulation overstates the observed effect by 0.7%.

**Figure C.28:** *Close to no association between document length and similarity in the pretreatment period*



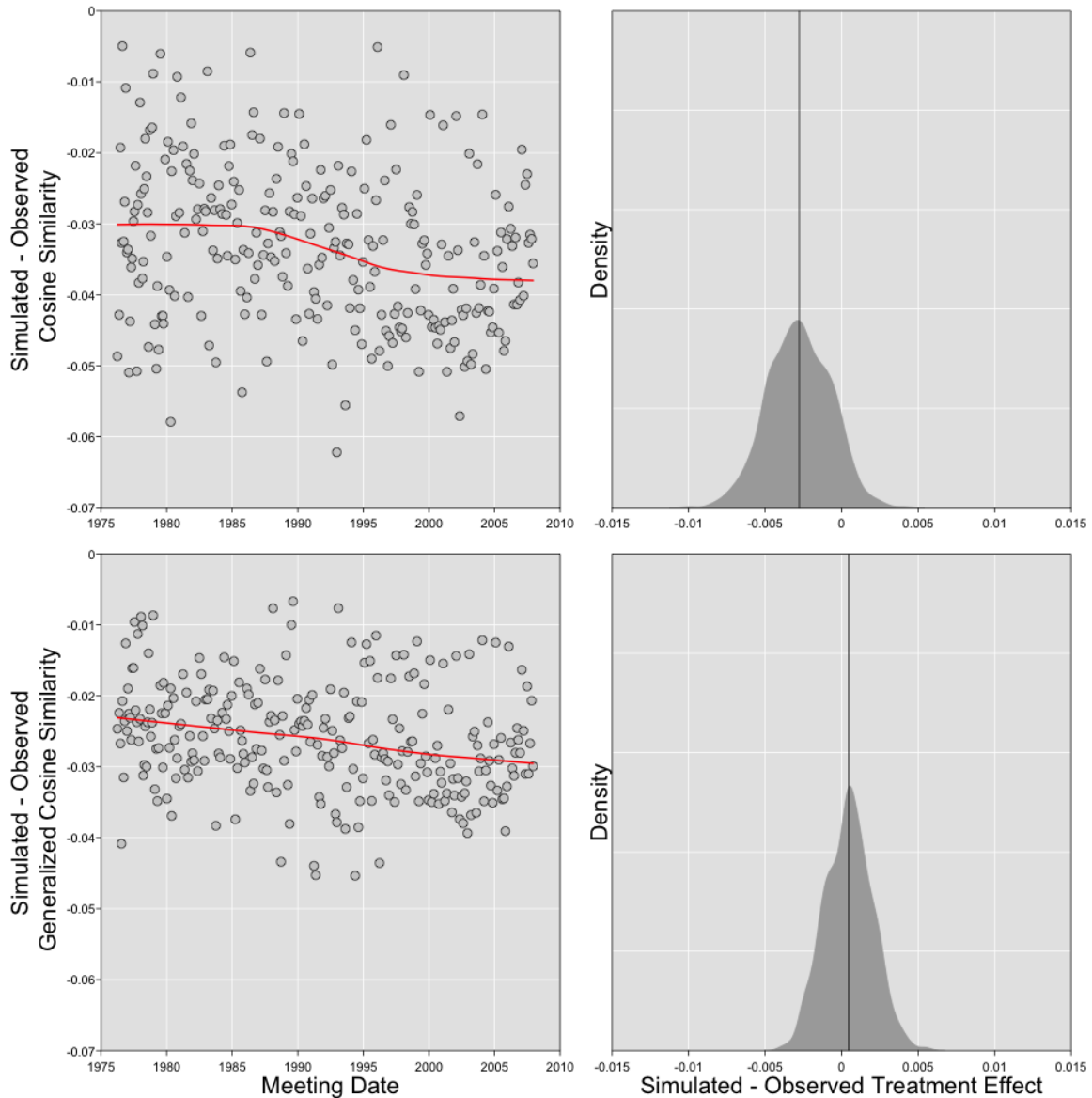
*Notes:* This figure shows the relationship between document length and similarity. Each point represents one meeting, with the  $x$ -axis representing the (standardized) geometric mean of the lengths of the public and private documents, and the  $y$ -axis representing the similarity of the public and private documents. The solid line is the best fit line (slope =  $-0.001$ ), and the dashed line is a locally weighted scatterplot smoothing.

## C.5 Theoretical Model

In this section, we develop a framework for considering the effects of transparency on deliberations. We maintain the intuition of Prat (2005) and Ottaviani and Sørensen (2001), where transparency can lead agents to pool their behavior around public signals. However, we build on those models by having the public's interest be of kind, not of degree.<sup>34</sup> We consider a representative FOMC member, abstracting from "conversation."

<sup>34</sup>That is to say, the public cares more about some topics, but does not have opinions on appropriate actions within each topic.

**Figure C.29:** Simulated change in similarity using documents generated from observed word proportions



*Notes:* We simulated 1000 iterations of both private and public documents. For each period, we took the true length and proportion of each word in each type of document, and used those as the underlying probabilities for a multinomial distributions, maintaining the true length of each document. Within each iteration, we then calculated the vector similarity of the counterfactual documents. In the left panels, we plot the gap between the mean simulated similarity and the similarity in the data. In the right panels, we show the distribution of the bias of estimated “treatment effects.”

We allow for two types of responses to transparency: a secular increase in effort, and a shift of effort across topics (e.g., Holmstrom, 1979; Holmstrom and Milgrom, 1991). A change in language due to combinations of the two will have ambiguous effects on welfare. The welfare effects of transparency, therefore, will be difficult to sign from just observing a change in language.

In each period, the FOMC undertakes effort in order to be able to say more about each topic in the meetings. The public cares about having high-effort FOMC members, but also may care differentially about effort placed on certain topics (such as inflation, as in Lucas Jr., 2000). The FOMC's utility comes from spending more effort discussing topics which ex-post were important, and potentially from appearing high-effort to the public.<sup>35</sup> In the public statements, the FOMC similarly expends effort to discuss each topic in more detail, and earns credit for spending more time on more important (to the public) topics.

The framework confirms the intuition that if the FOMC puts weight on public opinion, transparency will lead the FOMC to adjust its speech. In particular, the change of the cosine similarity between the deliberations and public statements captures will increase captures the relative amount that the FOMC cares about public opinion.

### Framework of FOMC Effort

The FOMC benefits from discussing more important topics more. For each of the  $I$  topics,  $\gamma_i \in (0,1)$  is the measure of the ex-post importance of topic  $i$ , with  $\mathbb{E}(\gamma_i) = p_i$ .  $\tau \in [0,1]$  indicates the level of transparency.<sup>36</sup>  $e_i$  denotes the amount of costly effort that the FOMC exerts on the topic: in order to discuss a topic in depth, more costly preparation is needed. The public judges the FOMC member through a weighted average of the total effort,

---

<sup>35</sup>In our interviews with central bankers, one concern with transparency was that it might particularly affect those who are looking to be written about "in the history books."

<sup>36</sup> $\tau$  is not necessarily equal to 0 before the policy reform—even though there were no public transcripts, some information about the deliberations may still have been made public, as in (Friedman and Schwartz, 1963).

where each topic has weight weights  $\sigma_i$ , and  $\sum \sigma_i = 1$ .  $\beta$ ,  $\alpha$ , and  $c$  mediate the amount that the FOMC member cares, respectively, about having put forth effort on important topics, public opinion, and exerting effort. The utility function from the deliberations is therefore:

$$\sum_i \beta \gamma_i e_i + \alpha \tau \sigma_i e_i - \frac{1}{2c} \cdot e_i^2.$$

The return to the press releases is similar, but with two differences. The first is that the press releases are, tautologically, fully transparent. Second, the only return to the press release is through public opinion—the FOMC gains no extra utility from discussing ex-post important topics. Utility from their press releases is therefore  $\sum_i \sigma_i m_i - \frac{1}{2d} \cdot m_i^2$ , where  $m_i$  is the effort used to publicly speak about issue  $i$ .

The optimal effort for each topic in the deliberations is  $e_i^*(\tau) = c(\beta p_i + \alpha \tau \sigma_i)$ . In the press release, optional effort is  $m_i^* = d \sigma_i$ . To a first order,<sup>37</sup> the growth in similarity from  $\tau_0$  to  $\tau_1$  is

$$\widehat{CS} = \sum_i \frac{\alpha (\tau_1 - \tau_0)}{\beta} \left( \frac{\sigma_i^2}{\sigma' \mathbf{p}} - \frac{\sigma_i p_i}{\mathbf{p}' \mathbf{p}} \right).$$

This implies that if we had a measure of the increase in similarity ( $\tau_1 - \tau_0$ ), we would be able to calculate  $\frac{\alpha}{\beta}$  given a change in topic usage. However, we do not observe topics, we observe word choices. The generalized cosine similarity approach can connect the two.

### From topics to words

In this subsection, we show that, under reasonable assumptions, there will be an increase in the cosine similarity of the observed language if and only if there is an increase in the similarity of the underlying topics. First, suppose that there are is some one-to-one and linear function  $f(\cdot) : \mathbb{R}^I \rightarrow \mathbb{R}^D$  which determines language choice as a function of topics. Second, suppose that the effort placed on each topic can be expressed as some linear combination of  $\sigma$  and  $\mathbf{p}$ . If  $CS(b\mathbf{p} + a_d \sigma, a_p \sigma) < CS(\tilde{b}\mathbf{p} + \tilde{a}_d \sigma, a_p \sigma)$ , then, by linearity

<sup>37</sup>Following Equation 3.1.



and monotonicity,  $CS(f(b\mathbf{p} + a_d\boldsymbol{\sigma}), f(a_p\boldsymbol{\sigma})) < CS(f(\tilde{b}\mathbf{p} + \tilde{a}_d\boldsymbol{\sigma}), f(a_p\boldsymbol{\sigma}))$ . As a result, a test of the the effect of transparency on deliberations is the effect of transparency on the cosine similarity of the words used in the public and private documents.

A way to generate  $f(\cdot)$  is through a topic to word matrix  $T$  where  $t_{ij}$  is the number of times word  $i$  is said for one “unit” of topic  $j$ , and with linearly independent columns. From word vectors  $w_0^p, w_1^p$ , and  $w^m$ , there are unique vectors  $\alpha_0^p, \alpha_1^p, \alpha^m$  such that  $w_0^p = T\alpha_0^p$ ,  $w_1^p = T\alpha_1^p$ ,  $w^m = T\alpha^m$ .

If we use  $(T(TT')^{-1}(TT')^{-1}T')$  as a weight matrix for the word vectors, we will find an increase in the generalized similarity of the word vectors if and only if there is an increase in similarity of the topic vectors. However, this will not satisfy all of the properties of generalized cosine similarity, since it is not positive definite, so we leave a full micro-foundation of semantic similarity measures for future research.

## References

- Abadie, Alberto. 2005. "Semiparametric Difference-in-Differences Estimators." *Review of Economic Studies* 72(1):1–19. 8
- Abadie, Alberto, Alexis Diamond and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105(490):493–505. 8, 11, 14, 17, 20, 23, 38, 40, 49, 55, 56
- Abadie, Alberto, Alexis Diamond and Jens Hainmueller. 2015. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59(2):495–510. 8, 14, 17, 20, 23, 38, 40, 49, 52, 55, 56
- Abadie, Alberto and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93(1):113–132. 14, 52, 55
- Acemoglu, Daron, Simon Johnson, Amir Kermani, James Kwak and Todd Mitton. 2016. "The value of connections in turbulent times: Evidence from the United States." *Journal of Financial Economics* 121(2):368–391. 56
- Acosta, Miguel. 2015. "FOMC Responses to Calls for Transparency." *Working Paper* . 93
- Allen, David and Matthew Connelly. 2015. Diplomatic History After the Big Bang: Using Computational Methods to Explore the Infinite Archive. In *Explaining the History of American Foreign Relations*, ed. Frank Costigliola and Michael Hogan. 3rd ed. New York: Cambridge University Press. 63
- Angeletos, George-Marios and Alessandro Pavan. 2004. "Transparency of Information and Coordination in Economies with Investment Complementarities." *The American Economic Review* 94(2):91–98. 65
- Angrist, Joshua D. 2014. "The perils of peer effects." *Labour Economics* 30:98–108. 94
- Angrist, Joshua D. and Jörn-Steffan Pischke. 2009. *Mostly Harmless Econometrics: An Empiricists Companion*. Princeton University Press. 73
- Arena, Philip and Scott Wolford. 2012. "Arms, Intelligence, and War." *International Studies Quarterly* 56(2):351–365. 52

- Atkinson, Anthony B. 1970. "On the Measurement of Inequality." *Journal of Economic Theory* 2(3):244–263. 100
- Auerbach, Robert D. 2009. *Deception and Abuse at the Fed: Henry B. Gonzalez Battles Alan Greenspan's Bank*. University of Texas Press. 96
- Auerbach, Robert D. 2011. "Audit the Fed: Dodd-Frank, QE 3, and Federal Reserve Transparency." *Testimony before Subcommittee on Domestic Monetary Policy and Technology, Committee on Financial Services, U.S. House of Representatives* . 90
- Barber, Brad M. and John D. Lyon. 1997. "Detecting long-run abnormal stock returns: The empirical power and specification of test statistics." *Journal of Financial Economics* 43:341–372. 19
- Barro, Robert J. 1981. "Output Effects of Government Purchases." *Journal of Political Economy* 89(6):1086–1121. 53
- Belasco, Amy. 2014. "The Cost of Iraq, Afghanistan, and Other Global War on Terror Operations Since 9/11." *Congressional Research Service Report 7-5700:1–106*. 10, 52, 129
- Bernanke, Ben S. and Kenneth N. Kuttner. 2005. "What Explains the Stock Market's Reaction to Federal Reserve Policy?" *The Journal of Finance* 60(3):1221–1258. 94
- Black, John, Nigar Hashimzade and Gareth Myles. 2009. *Oxford Dictionary of Economics*. Oxford University Press.
- Blei, David M. 2012. "Probabilistic Topic Models." *Communications of the ACM* 55(4):77–84. 150
- Blei, David M. 2013. "Stochastic Variational Inference." *Journal of Machine Learning Research* 14:1303–1347. 154
- Blei, David M., Andrew Y. Ng and Michael I Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 2(4-5):993—1022. 89, 104, 150, 154
- Bloom, Nicholas, Mark Schankerman and John Van Reenen. 2013. "Identifying Technology Spillovers and Product Market Rivalry." *Econometrica* 81(4):1347–1393. 89, 98, 128
- Boesler, Matthew and Liz McCormick. 2016. "QE's Cost: Fed Exit May Hit Economy Faster Than in Past Cycles." *Bloomberg News* . 10
- Bohn, Sarah, Magnus Lofstrom and Steven Raphael. 2014. "Did the 2007 Legal Arizona Workers Act Reduce the State's Unauthorized Immigrant Population?" *The Review of Economics and Statistics* 96(2):258–269. 55
- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45(1):5–32. 62, 77
- Brodersen, Kay H., Fabian Gallusser, Jim Koehler, Nicolas Remy and Steven L. Scott. 2015. "Inferring Causal Impact Using Bayesian Structural Time-Series Models." *The Annals of Applied Statistics* 9(1):247–274. 8, 14, 15, 18, 20, 55

- Calonico, Sebastian, Matias Cattaneo and Rocío Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82(6):2295–2326. 133, 135
- Campbell, John Y., Andrew W. Lo and A. Craig MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton University Press. 14, 22, 24, 30
- Caplan, B. 2002. "How does war shock the economy." *Journal of International Money and Finance* 21:145–162. 53
- Cavallo, Eduardo, Sebastian Galiani, Ilan Noy and Juan Pantano. 2013. "Catastrophic Natural Disasters and Economic Growth." *The Review of Economics and Statistics* 95(5):1549–1562. 55
- Chang, Jonathan and David M. Blei. 2009. "Relational Topic Models for Document Networks." *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics* pp. 1–8. 154, 157
- Chang, Jonathan and David M. Blei. 2010. "Hierarchical Relational Models for Document Networks." *The Annals of Applied Statistics* 4(1):124–150. 154, 155, 157
- Claessens, Stijn, Erik Feijen and Luc Laeven. 2008. "Political Connections and Preferential Access to Finance: The Role of Campaign Contributions." *Journal of Financial Economics* 88(3):554–580. 14
- Clinton, Joshua D. and David E. Lewis. 2008. "Expert Opinion, Agency Characteristics, and Agency Preferences." *Political Analysis* 16(1):3–20. 5
- Colaresi, Michael. 2014. *Democracy Declassified: The Secrecy Dilemma in National Security*. Oxford: Oxford University Press. 58
- Collier, Paul. 1999. "On the Economic Consequences of Civil War." *Oxford Economic Papers* 51:168–183. 52
- Cornand, Camille and Frank Heinemann. 2008. "Optimal Degree of Public Information Dissemination." *The Economic Journal* 118:718–742. 65
- Cox, David R. 1958. *Planning of Experiments*. Wiley Publications in Statistics. 75
- Cukierman, Alex. 2001. Accountability, credibility, transparency, and stabilization policy in the Eurosystem. In *The Impact of the EMU on Europe and Developing Countries*, ed. C. Wyplosz. Oxford University Press. 65
- Danker, Deborah J. and Matthew M. Luecke. 2005. "Background on FOMC Meeting Minutes." *Federal Reserve Bulletin* Spring:175–179. 95
- Davis, Lucas W. 2008. "The Effect of Driving Restrictions on Air Quality in Mexico City." *Journal of Political Economy*, 116(1):38–81. 133

- Davis, Steven J. and John Haltiwanger. 1992. "Gross Job Creation, Gross Job Destruction, and Employment Reallocation." *The Quarterly Journal of Economics* 107(3):819–863. 103
- de Figueiredo, John M. and Brian Kelleher Richter. 2014. "Advancing the Empirical Research on Lobbying." *Annual Review of Political Science* 17:163–185. 7
- Deger, Saadet and Somnath Sen. 1995. Military expenditure and developing countries. In *Handbook of Defense Economics*, ed. K. Hartley and T. Sandler. Vol. 1 Elsevier Science B. V. chapter 11, pp. 272–307. 52
- Drakos, Konstantinos. 2004. "Terrorism-induced structural shifts in financial risk: airline stocks in the aftermath of the September 11th terror attacks." *European Journal of Political Economy* 20(2):435–446. 28
- Dranove, David, Daniel Kessler, Mark McClellan and Mark Satterthwaite. 2003. "Is More Information Better? The Effects of 'Report Cards' on Health Care Providers." *Journal of Political Economy* 111(3):555–588. 93
- Drutman, Lee. 2015. *The Business of America is Lobbying: How Corporations Became Politicized and Politics Became More Corporate*. Oxford University Press. 9
- Durbin, J. and S. J. Koopman. 2002. "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis." *Biometrika* 89(3):603–615. 15, 18, 20
- Easterly, William. 2003. "Can Foreign Aid Buy Growth?" *The Journal of Economic Perspectives* 17(3):23–48. 10
- Eldor, Rafi and Rafi Melnick. 2004. "Financial markets and terrorism." *European Journal of Political Economy* 20(2):367–386. 26
- Fama, Eugene F. 1980. "Agency Problems and the Theory of the Firm." *The Journal of Political Economy* 88(2):288–307. 93
- Fama, Eugene F. 1998. "Market efficiency, long-term returns, and behavioral finance." *Journal of Financial Economics* 49(3):283–306. 19
- Fearon, James. 1994a. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88:577–92. 59, 61, 64
- Fearon, James. 1994b. "Signaling Versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model." *Journal of Conflict Resolution* 38(2):236–69. 59
- Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49(3):379–414. 52
- Fey, Mark and Kristopher Ramsay. 2011. "Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict." *American Journal of Political Science* 55(1):149–169. 59

- Fisman, Raymond. 2001. "Estimating the Value of Political Connections." *American Economic Review* 91(4):1095–1102. 14, 34
- Fligstein, Neil, Jonah Stuart Brundage and Michael Schultz. 2014. "Why the Federal Reserve Failed to See the Financial Crisis of 2008: The Role of "Macroeconomics" as a Sensemaking and Cultural Frame." *IRLE Working Paper No. 111–14* pp. 1–65. 94
- Fowler, Anthony. 2013. "Electoral and Policy Consequences of Voter Turnout: Evidence from Compulsory Voting in Australia." *Quarterly Journal of Political Science* 8(159–182). 56
- Fox, J. Ronald. 2011. *Defense Acquisition Reform: An Elusive Goal*. Washington, D.C.: Center of Military History: United States Army. 5
- Frankel, David M and Oscar Volij. 2011. "Measuring School Segregation." *Journal of Economic Theory* 146(1):1–38. 98, 100
- Frieden, Jeffry and David Lake. 2005. "International Relations as a Social Science: Rigor and Relevance." *Annals of the American Academy of Political and Social Science* 600(1):136–156. 58
- Friedman, Milton and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867-1960*. Princeton University Press for NBER. 161
- Friedman, Milton and Anna Schwartz. 1993. "A Tale of Fed Transcripts." *The Wall Street Journal* . 96
- Gaikwad, Nikhar. 2013. "Presidential Prospects, Political Support, and Stock Market Performance." *Quarterly Journal of Political Science* 8:1–14. 14
- Garten, Jeffrey E. 1997. "Business and Foreign Policy." *Foreign Affairs* May/June. 5
- Gelman, Andrew and Guido Imbens. 2014. "Why High-Order Polynomials Should Not be Used in Regression Discontinuity Designs." *NBER Working Paper No. 20405* . 133
- Gentzkow, Matthew and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica* 78(1):35–71. 89
- Gentzkow, Matthew, Jesse Shapiro and Matt Taddy. 2015. "Measuring Polarization in High-Dimensional Data: Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech." *Working Paper* . 103, 158
- George, Edward I. and Robert E. McCulloch. 1993. "Variable Selection Via Gibbs Sampling." *Journal of the American Statistical Association* 88(423):881–889. 20, 131
- Geraats, Petra M. 2002. "Central Bank Transparency." *The Economic Journal* 112(483):F532–F565. 94

- Ghosn, Faten, Glenn Palmer and Stuart Bremer. 2004. "The MID3 Data Set, 1993–2001: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 21(2):133–154. 60
- Gill, Michael and Arthur Spirling. 2015. "Estimating the Severity of the WikiLeaks U.S. Diplomatic Cables Disclosure." *Political Analysis* 23(2):299–305. 67
- Goemans, Hein. 2000. *War and Punishment: The Causes of War Termination and the First World War*. Princeton, NJ: Princeton University Press. 59
- Griffiths, Thomas L. and Mark Steyvers. 2004. "Finding scientific topics." *Proceedings of the National Academy of Sciences* 101(suppl 1):5228–5235. 89
- Grimmer, Justin and Brandon M. Stewart. 2013a. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* pp. 1–31. 89, 90
- Grimmer, Justin and Brandon Stewart. 2013b. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–27. 62, 72
- Hansen, Stephen, Michael McMahon and Andrea Prat. 2014. "Transparency and Deliberation within the FOMC: a Computational Linguistics Approach." *Working Paper* . 93, 116
- Hastie, Trevor, Robert Tibshirani and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer. 143
- Heilmann, Kilian. 2015. "Does political conflict hurt trade? Evidence from consumer boycotts." *Journal of International Economics* 99:179–191. 56
- Herron, Michael C. 2000. "Estimating the Economic Impact of Political Party Competition in the 1992 British Election." *American Journal of Political Science* 44(2):326–337. 14
- Ho, Tin. 1998. "The Random Subspace Method for Constructing Decision Forests." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(8):832–844. 143
- Hoberg, Gerard and Gordon M. Phillips. 2010. "Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis." *Review of Financial Studies* 23(10):3773–3811. 91
- Hoberg, Gerard and Gordon M. Phillips. 2016. "Text-Based Network Industries and Endogenous Product Differentiation." *Journal of Political Economy* Forthcoming. 91
- Hoffman, Matthew D., David M. Blei, Chong Wang and John Paisley. 2013. "Stochastic Variational Inference." *Journal of Machine Learning Research* 14:1303–1347. 150
- Holland, P.W. 1986. "Statistics and causal inference." *Journal of the American Statistical Association* 81:945–968. 13

- Holm, Sture. 1979. "A simple sequentially rejective multiple test procedure." *Scandinavian Journal of Statistics* 6(2):65–70. 79
- Holmstrom, Bengt. 1979. "Moral Hazard and Observability." *The Bell Journal of Economics* 10(1):74–91. 93, 161
- Holmstrom, Bengt and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7(Papers from the Conference on the New Science of Organization):24–52. 161
- Iacus, Stefano M., Gary King and Giuseppe Porro. 2011. "Multivariate Matching Methods That Are Monotonic Imbalance Bounding." *Journal of the American Statistical Association* 106(493):345–361. 75
- Imai, Kosuke, Gary King and Elizabeth A. Stuart. 2008. "Misunderstandings between experimentalists and observationalists about causal inference." *Journal of the Royal Statistical Society, Series A* 171(2):481–502. 75
- Ishwaran, Hemant and J. Sunil Rao. 2005. "Spike and Slab Variable Selection: Frequentist and Bayesian Strategies." *The Annals of Statistics* 33(2):730–773. 20, 131
- Jacobs, Lawrence R. and Benjamin I. Page. 2005. "Who Influences U.S. Foreign Policy?" *American Political Science Review* 99(1):107–123. 5
- Jayachandran, Seema. 2006. "The Jeffords Effect." *Journal of Law and Economics* 49(2):397–425. 14
- Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton, N.J.: Princeton University Press. 60
- Joe, Harry. 2006. "Generating random correlation matrices based on partial correlations." *Journal of Multivariate Analysis* 97(10):2177–2189. 45, 46
- Kagan, Kimberly. 2010. *The Surge: A Military History*. Encounter Books. 26
- Kang, Karam. 2015. "Policy Influence and Private Returns from Lobbying in the Energy Sector." *The Review of Economic Studies* Forthcoming. 91
- Keane, Jack and Frederick W. Kagan. 2006. "The Right Type of 'Surge'." *The Washington Post* (December 27, 2006). 26
- Keele, Luke, Neil Malhotra and Colin H. McCubbins. 2013. "Do Term Limits Restrain State Fiscal Policy? Approaches for Causal Inference in Assessing the Effects of Legislative Institutions." *Legislative Studies Quarterly* 38(3):291–326. 55
- Keohane, Robert and Joseph Nye. 1977. *Power and Interdependence: World Politics in Transition*. Boston: Little, Brown. 59, 61, 80



- Kleven, Henrik Jacobsen, Camille Landais and Emmanuel Saez. 2013. "Taxation and International Migration of Superstars: Evidence from the European Football Market." *American Economic Review* 103(5):1892–1924. 56
- Koenker, Roger and Kevin F. Hallock. 2001. "Quantile Regression." *Journal of Economic Perspectives* 15(4):143–156. 34
- Kothari, S. and J. B. Warner. 2008. Econometrics of event studies. In *Handbook of Corporate Finance: Empirical Corporate Finance*. Vol. 1 Elsevier/North Holland. 8, 19
- Kurizaki, Shuhei. 2007. "Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy." *American Political Science Review* 101(3). 59, 83, 85, 87
- Lai, Tze Leung and Haipeng Xing. 2010. *Statistical Models and Methods for Financial Markets*. Springer. 22
- Lake, David. 2010. "Two Cheers for Bargaining Theory: Assessing Rationalist Explanations of the Iraq War." *International Security* 35(3):7–52. 58
- Lee, David S. and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48:281–355. 133
- Lesk, Michael. 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on Systems documentation*. ACM New York, NY: pp. 24–26. 93
- Lessard, Donald R. 1974. "World, National, and Industry Factors in Equity Returns." *The Journal of Finance* 29(2):379–391. 22
- Lewellen, Jonathan. 2002. "Momentum and Autocorrelation in Stock Returns." *The Review of Financial Studies* 15(2):533–563. 14
- Liben-Nowell, David and Jon Kleinberg. 2003. The Link Prediction Problem for Social Networks. In *The Twelfth Annual ACM International Conference on Information and Knowledge Management*. pp. 556–559. 154
- Lo, Andrew W. and A. Craig MacKinlay. 1990. "When are Contrarian Profits Due to Stock Market Overreaction?" *The Review of Financial Studies* 3(2):175–205. 14
- Lucas Jr., Robert E. 2000. "Inflation and Eelfare." *Econometrica* 68(2):247–274. 161
- MacKinlay, A. Craig. 1997. "Event Studies in Economics and Finance." *Journal of Economic Literature* 35(1):13–39. 14, 24
- Manning, Christopher D., Prabhakar Raghavan and Hinrich Schütze. 2008. *An Introduction to Information Retrieval*. Cambridge University Press. 150
- Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60(3):531–542. 94

- McHale, Michael. 1998. A comparison of WordNet and Roget's taxonomy for measuring semantic similarity. In *Proceedings COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems*. ACM Montreal, Canada: pp. 115–120. 93
- Meade, Ellen E. and Daniel L. Thornton. 2012. "The Phillips Curve and US Monetary Policy: What the FOMC Transcripts Tell Us." *Oxford Economic Papers* 64(2):197–216. 94
- Meade, Ellen E. and David Stasavage. 2008. "Publicity of Debate and the Incentive to Dissent: Evidence from the US Federal Reserve." *The Economic Journal* 118(2001):695–717. 93
- Mendelson, Edward. 2008. "ABBYY FineReader."
- Meyer, Bruce D. 1995. "Natural and Quasi-Experiments in Economics." *Journal of Business and Economic Statistics* 13(3):151–161. 17
- Miller, George A. 1995. "WordNet: a lexical database for English." *Communications of the ACM* 38(11):39–41. 93
- Miller, George A, Richard Beckwith, Christiane Fellbaum, Derek Gross and Katherine J Miller. 1990. "Introduction to wordnet: An on-line lexical database." *International Journal of Lexicography* 3(4):235–244. 93
- Mintz, Alex. 1989. "Guns Versus Butter: A Disaggregated Analysis." *American Political Science Review* 83(4):1285–1293. 5, 52
- Mintz, Alex and Chi Huang. 1992. "Defense expenditures, economic growth, and the 'peace dividend'." *American Political Science Review* 84(4):1283–1293. 52, 53
- Mitchell, T.J. and J.J. Beauchamp. 1988. "Bayesian Variable Selection in Linear Regression." *Journal of the American Statistical Association* 83(404):1023–1032. 20, 131
- Navigli, Roberto. 2009. "Word sense disambiguation: A survey." *ACM Computing Surveys (CSUR)* 41(2):1–69. 89
- Newey, Whitney K. and Kenneth D. West. 1994. "Automatic Lag Selection in Covariance Matrix Estimation." *The Review of Economic Studies* 61(4):631–653. 108
- Northrop Grumman Corporation. 2014. *10-K Report*. Retrieved from EDGAR System database. (CIK No. 0001133421). 6
- Ottaviani, Marco and Peter Sørensen. 2001. "Information aggregation in debate: who should speak first?" *Journal of Public Economics* 81(3):393–421. 159
- Park, Trevor and George Casella. 2008. "The Bayesian Lasso." *Journal of the American Statistical Association* 103(482):681–686. 20
- Partell, Peter and Glenn Palmer. 1999. "Audience Costs and Interstate Crises: An Empirical Assessment of Fearon's Model of Dispute Outcomes." *International Studies Quarterly* 43(2):389–405. 59

- Pattison, James. 2014. *The Morality of Private War: The Challenge of Private Military and Security Companies*. Oxford University Press. 5
- Percy, Sarah. 2007. *Mercenaries: The History of a Norm in International Relations*. Oxford University Press. 5
- Peters, Heidi M., Moshe Schwartz and Lawrence Kapp. 2015. "Department of Defense Contractor and Troop Levels in Iraq and Afghanistan: 2007–2015." *Congressional Research Service Report (7-5700, R44116)* pp. 1–9. 30
- Phelan, Christopher and Alberto Trejos. 2000. "The aggregate effects of sectoral reallocations." *Journal of Monetary Economics* 45(2):249–268. 53
- Phillips, Brian J. 2015. "Civil war, spillover, and neighbors' military spending." *Conflict Management and Peace Science* 32(4):425–442. 52
- Pinotti, Paolo. 2015. "The Economic Costs of Organised Crime: Evidence from Southern Italy." *The Economic Journal* 125(Feature Issue):F203–F232. 56
- Porter, Martin F. 1980. "An Algorithm for Suffix Stripping." *Program* 14(3):130–137. 72, 97
- Potter, Philip and Matthew Baum. 2014. "Looking for Audience Costs in all the Wrong Places: Electoral Institutions, Media Access, and Democratic Constraint." *The Journal of Politics* 76:167–181. 59
- Powell, Robert. 1993. "Guns, Butter, and Anarchy." *American Political Science Review* 87(1):115–132. 52
- Powell, Robert. 1999. *In the Shadow of Power*,. Princeton, NJ: Princeton University Press. 58
- Powell, Robert. 2002. "Bargaining Theory and International Conflict." *Annual Review of Political Science* 5:1–30. 58, 63
- Prat, Andrea. 2005. "The Wrong Kind of Transparency." *American Economic Review* 95(3):862–877. 93, 159
- Quinn, Kevin, Burt Monroe, Michael Colaresi, Michael H. Crespin and Dragomir Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54:209–228. 150
- Ram, Rati. 1995. Defense Expenditure and Economic Growth. In *Handbook of Defense Economics*, ed. K. Hartley and T. Sandler. Vol. 1 Elsevier Science B. V. chapter 10, pp. 251–274. 52, 53
- Ramey, Valerie A. 2011. "Identifying Government Spending Shocks: It's All in the Timing." *The Quarterly Journal of Economics* 126(1):1–50. 5, 53
- Ramey, Valerie A. and Matthew D. Shapiro. 1999. "Costly Capital Reallocation and the Effects of Government Spending." *NBER Working Paper No. 6283* pp. 1–64. 5, 52, 53

- Ramsay, Kristopher. 2008. "Settling It on the Field: Battlefield Events and War Termination." *Journal of Conflict Resolution* 52(6):850–879. 59
- Ramsay, Kristopher. 2011. "Cheap Talk Diplomacy, Voluntary Negotiations, and Variable Bargaining Power." *International Studies Quarterly* 55:1003–1023. 59
- Raytheon Company. 2014. *10-K Report*. Retrieved from EDGAR System database. (CIK No. 0001047122). 6
- Reed, William. 2003. "Information and Economic Interdependence." *Journal of Conflict Resolution* 47(1):54–71. 59
- Reich, Michael. 1972. "Does the US Economy Require Military Spending?" *American Economic Review* 62(1):296–303. 5, 53
- Reiter, Dan. 2009. *How Wars End*. Princeton: Princeton University Press. 59
- Renshon, Jonathan, Jooa Julia Lee and Dustin Tingley. 2015. "Physiological Arousal and Political Beliefs." *Political Psychology* 36(5):569–585. 60
- Resnik, Philip. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *14th International Joint Conference on Artificial Intelligence, IJCAI*, ed. C.S. Mellish. Morgan Kaufmann Publishers Inc. pp. 448–453. 93
- Roberts, Brian E. 1990. "Political Institutions, Policy Expectations, and the 1980 Election: A Financial Market Perspective." *American Journal of Political Science* 34(2):289–310. 14
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, 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. 89
- Romer, Christina D. and David Romer. 2004. "A new measure of monetary shocks: Derivation and implications." *The American Economic Review* 94(4):1055–1084. 94
- Rosenbaum, Paul R. 2007. "Interference Between Units in Randomized Experiments." *Journal of the American Statistical Association* 102(477):191–200. 17
- Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1):41–55. 17
- Rotemberg, Julio. 2013. "Penitence after accusations of error: 100 Years of Monetary Policy at the US Federal Reserve." *Journal of Economic Perspectives* 26(4). 94
- Rubin, Donald B. 1976. "Multivariate Matching Methods That Are Equal Percent Bias Reducing, I: Some Examples." *Biometrics* 32(1):109–120. 75
- Rudolph, Kara E., Elizabeth A. Stuart, Jon S. Vernick and Daniel W. Webster. 2015. "Association between Connecticut's permit-to-purchase handgun law and homicides." *American Journal of Public Health* 105(8):e49–354. 56

- Russett, Bruce. 1976. "Defense Expenditures and National Well-Being." *American Political Science Review* 76(4):767–777. 5
- Salton, Gerard and Michael J. McGill. 1986. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc. 91
- Sandler, Todd. 2014. "The analytical study of terrorism: Taking stock." *Journal of Peace Research* 51(2):257–271. 52
- Sartori, Anne. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 56(1):121–149. 59, 61, 83, 85, 87
- Sartori, Anne. 2005. *Deterrence by Diplomacy*. Princeton: Princeton University Press. 59
- Schneider, Gerald and Vera E. Troeger. 2006. "War and the World Economy: Stock Market Reactions to International Conflicts." *Journal of Conflict Resolution* 50:623–645. 53
- Schonhardt-Bailey, Cheryl. 2013. *Deliberating American Monetary Policy*. MIT Press. 94
- Schultz, Kenneth. 2001. "Looking for Audience Costs." *Journal of Conflict Resolution* 45:32–60. 63
- Scott, Steven L. and Hal R. Varian. 2014. "Predicting the present with Bayesian structural time series." *International Journal of Mathematical Modeling and Optimization* 5:4–23. 14, 15, 20
- Shapiro, Jacob and David Siegel. 2010. "Is this Paper Dangerous? Balancing Secrecy and Openness in Counterterrorism." *Security Studies* 19(1):66–98. 60
- Signorino, Curtis. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93(2):279–297. 59
- Singer, P.W. 2008. *Corporate Warriors: The Rise of the Privatized Military Industry*. Cornell University Press. 5
- Slantchev, Branislav. 2004. "How Initiators End Their Wars: The Duration of Warfare and the Terms of Peace." *American Journal of Political Science* 48(4):813–829. 59
- Slantchev, Branislav. 2012. "Audience Cost Theory and Its Critics." *Security Studies* 21:376–382. 86
- Smith, Alistair and Allan Stam. 2004. "Bargaining and the Nature of War." *Journal of Conflict Resolution* 48(6). 59
- Smith, Ron. 2009. *Military Economics: The Interaction of Power and Money*. Palgrave Macmillan. 5
- Snowberg, Erik, Justin Wolfers and Eric Zitzewitz. 2011. How Prediction Markets can Save Event Studies. In *Prediction Markets*, ed. Lieghton Vaughn Williams. Routledge. 25

- Snyder, Jack and Erica D. Borghard. 2011. "The Cost of Empty Threats: A Penny, Not a Pound." *American Political Science Review* 105(3). 63
- Stiglitz, Joseph E. and Linda J. Bilmes. 2008. *The Three Trillion Dollar War: The True Cost of the Iraq Conflict*. W.W. Norton. 52
- Stone, Philip J. and Earl B. Hunt. 1963. A Computer Approach to Content Analysis: Studies Using the General Inquirer System. In *Proceedings of the May 21-23, 1963, Spring Joint Computer Conference*. AFIPS '63 (Spring) New York, NY, USA: ACM pp. 241-256. 104
- Street, Alex, Thomas A. Murray, John Blitzer and Rajan S. Patel. 2015. "Estimating Voter Registration Deadline Effects with Web Search Data." *Political Analysis* 23(2):225-241. 14
- Stuart, Elizabeth A. 2010. "Matching methods for causal inference: A review and a look forward." *Statistical Science* 25(1):1-21. 75
- Swanson, Eric T. 2006. "Have Increases in Federal Reserve Transparency Improved Private Sector Interest Rate Forecasts?" *Journal of Money, Credit and Banking* 38(3):791-819. 94
- Tarar, Ahmer and Behar Leventoglu. 2009. "Public Commitment in Crisis Bargaining." *International Studies Quarterly* 53:817-839. 59
- Thucydides. 1910/431BCE. *The Peloponnesian War*. London: J. M. Dent. 85
- Tibshirani, Robert. 1994. "Regression Shrinkage and Selection Via the Lasso." *Journal of the Royal Statistical Society, Series B* 58:267-288. 62, 77, 143
- Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61:821-40. 63
- Trachtenberg, Marc. 2012. "Audience Costs: An Historical Analysis." *Security Studies* 21(1):3-42. 63
- Trager, Robert. 2010. "Diplomatic Calculus in Anarchy: How Communication Matters." *American Political Science Review* 104(2). 59, 86, 87
- Turney, Peter D. and Patrick Pantel. 2010. "From Frequency to Meaning: Vector Space Models of Semantics." *Journal of Artificial Intelligence Research* 37:141-188. 89
- Tversky, Amos. 1977. "Features of similarity." *Psychological Review* 84(3):327-352. 98
- Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28(2):3-28. 14, 55
- von Clausewitz, Carl. 1832/1989. *On War*. Princeton: Princeton University Press. 85
- Wallach, Hanna M., Iain Murray, Ruslan Salakhutdinov and David Mimno. 2009. Evaluation Methods for Topic Models. In *Proceedings of the 26th International Conference on Machine Learning*. 90

- Wallach, Hannah, David M. Mimno and Andrew McCallum. 2009. Rethinking LDA: Why priors matter. In *Neural Information Processing Systems*. NIPS. 90
- Waltz, Kenneth. 1979. *Theory of International Politics*. New York: McGraw-Hill. 59
- Weeks, Jessica. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62(1). 61, 63
- Weidmann, Nils. 2016. "A Closer Look at Reporting Bias in Conflict Event Data." *American Journal of Political Science* 60(1):206–218. 62
- Wendt, Alexander. 1999. *Social Theory of International Politics*. New York: Cambridge University Press. 59
- Werner, Suzanne. 1999. "The Precarious Nature of Peace: Resolving the Issues, Enforcing the Settlement, and Renegotiating the Terms." *American Journal of Political Science* 43(3):912–934. 59
- Whitten, Guy D. and Laron K. Williams. 2011. "Buttery Guns and Welfare Hawks: The Politics of Defense Spending in Advanced Industrial Democracies." *American Journal of Political Science* 55(1):117–134. 52
- Williams, Laron K. 2015. "Estimating the defense spending vote." *Electoral Studies* 39:243–255. 52
- Woodford, Michael. 2005. "Central Bank Communication and Policy Effectiveness." *NBER Working Paper No. 11898* . 65
- Woolley, John T and Joseph Gardner. 2009. Does Sunshine Reduce the Quality of Deliberation? The Case of the Federal Open Market Committee. In *Annual Meeting of the American Political Science Association*. 93